

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 19, 2003, 14:59:25 ; Search time 28 Seconds
(without alignments)
861.776 Million cell updates/sec

Title: US-09-903-410-36

Sequence: 1273
1 LRLRRFEENINVLISGGAANG.....RKADQIMRGYIKALEVLSE 251

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

PIR 73: *
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	934	73.4	259	2	F70420 conserved hypotnet
2	391.5	30.8	728	2	E83328 hypothetical prote
3	300.5	23.6	275	2	C83973 hypothetical prote
4	297.5	23.4	764	2	B82303 conserved hypotnet
5	276	21.7	293	2	AB3639 serine proteinase
6	274.5	21.6	300	2	A82017 probable lipoprote
7	273	21.4	312	2	AH2791 conserved hypotnet
8	273	21.4	314	2	AF3365 serine proteinase
9	273	21.4	315	2	H97570 hypothetical 34.4K
10	268.5	21.1	260	2	H69874 conserved hypotnet
11	268.5	21.1	297	2	A81002 conserved hypotnet
12	268.5	21.1	395	2	F82852 conserved hypotnet
13	258.5	20.3	314	2	G85703 hypothetical prote
14	258.5	20.3	314	2	H90845 hypothetical prote
15	257.5	20.2	301	2	AH0649 conserved hypotnet
16	257.5	20.2	314	2	B36871 probable membrane
17	253.5	19.9	1048	2	A70592 hypothetical prote
18	243	19.1	345	2	G83441 conserved hypotnet
19	242	19.0	1679	2	S49802 probable membrane
20	240	18.9	311	2	D71252 conserved hypotnet
21	235.5	18.5	610	2	B87518 conserved hypotnet
22	235	18.5	583	2	A70729 hypothetical prote
23	234	18.4	360	2	E70892 hypothetical prote
24	231.5	18.2	308	2	AE2661 conserved hypotnet
25	231.5	18.2	311	2	C97443 hypothetical prote
26	231	18.1	1065	2	A70797 hypothetical prote
27	223.5	17.6	291	2	F69959 hypothetical prote
28	208.5	16.4	1351	2	S44655 ZK170.4 protein -
29	199.5	15.7	296	2	E84000 hypothetical prote

30	189.5	14.9	860	2	T21745 hypothetical prote
31	162	12.7	283	2	H97198 probable phosphos
32	161	12.6	1316	2	T50444 hypothetical UPF00
33	153	12.0	254	2	G72343 conserved hypotnet
34	152.5	12.0	324	2	D70943 hypothetical prote
35	147.5	11.6	329	2	A87087 conserved hypotnet
36	147.5	11.6	598	2	E71657 hypothetical prote
37	142	11.2	356	2	E85663 hypothetical prote
38	142	11.2	356	2	G90803 hypothetical prote
39	142	11.2	687	2	T09051 hypothetical prote
40	139.5	11.0	288	2	T17654 probable aspartate
41	138.5	10.9	353	2	E87668 conserved hypotnet
42	138.5	10.9	749	2	S61643 probable membrane
43	134	10.5	801	2	T47774 hypothetical prote
44	134	10.5	825	2	T48431 hypothetical prote
45	133	10.4	356	2	F90978 hypothetical prote

ALIGNMENTS

RESULT 1

F70420

conserved hypothetical protein aq_1386 - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999

C:Accession: F70420

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.

V. Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: F70420

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-259 <AOP>

A:Cross-references: GB:A800737; NID:92983782; PIDN:AAC07357.1; PID:92983799; GB:AE0

A:Experimental source: strain VFS

C:Genetics:

A:Gene: aq_1386

Query Match	73.4%	Score 934	DB 2	Length 259
Best Local Similarity	71.5%	Pred. No. 9.9e-74		
Matches 178	Conservative 31	Mismatches 40	Indels 0	Gaps 0
QY	1	LRLRRFEENINVLISGGAANGIAHIGVLAINEIGIRVALSGVSAIVSVFASGYSPB	60	
DB	5	LKLRRFEENINVLISGGAANGIAHIGVLAINEIGIRVALSGVSAIVSVFASGYSPB	64	
QY	61	GMSILKRVNMLKFKFPLKGLIGWKAIRPLEVLPYRLEKLEIPYTCATDLXSG	120	
DB	65	EWLKLKRVNMLKFKFPLKGLIGWKAIRPLEVLPYRLEKLEIPYTCATDLXSG	124	
QY	121	RALVYSEGSILPALIGSCAIPGIFEPVEYKYNILVVDGIVNNLPVEPQESGIPYCVDV	180	
DB	125	KALVYGRDILIPVLLGSGSIPGIFEPVEYKYNILVVDGIVNNLPVEPQESGIPYCVDV	184	
QY	181	LPIERKQIKNIMLILASFLAVRSNKKRKECDLVYPELSEFPPLDVRKADQIMER	240	
DB	185	LPIERKQIKNIMLILASFLAVRSNKKRKECDLVYPELSEFPPLDVRKADQIMER	244	
QY	241	GYIKALEVL 249		
DB	245	GYESTLRIM 253		

RESULT 2

E83328

hypothetical protein PA339 [imported] - Pseudomonas aeruginosa (strain PA01).

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: E83328

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;

A;Authors: Foulger, D., Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galielich, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Huilo, P.; Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardini, A.;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maizumi, Y.; Matsuda, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parvo, V.; Pohl, T.M.; Portet-Berle, R.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scandellari, A.;Authors: Reich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, Y.; Sekowska, A.; Setbon, M.; Authors: Schleich, S.; Schroeter, R.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, K.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.;Authors: Yoshioka, H.P.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*

A;Reference number: A69580; PMID:98044033; PMID:9384377

A;Accession: H5984

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-260 <K>

A:Cross-references: GB:Z99111; GB:A009126; NID:G2633699; PIDN:CAB13377.1; PID:e1185C

A:Experimental source: strain 168

C:Genetics:

A:Gene: vlbk

Query Match	21.1%	Score	268.5;	DB 2;	Length	260;			
Best Local Similarity	28.2%;	Prod. No.	7.7e-16;						
Matches	68;	Conservative	54;	Mismatches	106;	Indels	13;	Gaps	4;
Qy	8	EINVLISGGAAGIAHIGVLKAINELGIRVRALSGVSAAGIVSFYASGYSPEGMFSLLK	67						
Db	5	KIGLALSGGGARGLAHIGVLSLKHQIIEVDMIAGSSMGALVGSFYAAGHD----	VATMK	60					
Qy	68	RVNWLKLPK-----FPPLKGLHIGNEKAIRPLEEVLPYRRLEKLEIETIYCNDLISGR	121						
Db	61	KV--AKAFKRLVADYTVPKGLPKGDRVRQLVHYATFGKPIBELQIPLGIVACDLQTE	118						
Qy	122	ALYLSSEGLIPALLGSCAIPGIEPEPVVEYKYNLLVDGGIVNNLPVFPFQESGIPVCV--DV	180						
Db	119	KIVPKGVSVDVRASISIGIIFPQLDGRLLVDGAVVDRIPVSVVKWDGADIIIASDV	178						
Qy	181	LPIEPEKDKINLHILARSFFLAVRSNKKFECDLVLPBELBEFTPLDVRKADQIMER	240						
Db	179	SVKRTETAVAHIVDMQMSDWILQNELVRHQTIAADIMIRPSLETISYSSSFANIEEISA	238						
Qy	241	G	241						
Db	239	G	239						

RESULT 11
A81002
conserved hypothetical protein NME2139 [imported] - Neisseria meningitidis (strain MC58)
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: A81002
R:Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.; Iri, H.; Qin, H.; Vamathchevan, J.; Gill, J.; Scariato, V.; Masignani, V.; Pizza, M. Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R. A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: A81002
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-297 <TEXT>
A:Cross-references: GB:AE002562; GB:AE002098; NID:G7227392; PIDN:AAF42447.1; PID:G7227392
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NME2139

Query Match 21.1%; Score 268.5; DB 2; Length 297;
Best Local Similarity 31.4%; Pred. No. 9.1e-16;
Matches 70; Conservative 41; Mismatches 97; Indels 15; Gaps 3;
QY 9 INLVLSGGAAGIAHIGVLKINELGTRVALSGVSGAIVSVFYASGYSPEGMFLIKR 68

[illegible]

RESULT 9

H97570
Hypothetical 34.4 K protein in hnr-puru intergenic region [imported] - Agrobacterium tumefaciens
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: H97570
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; PMID:11743194
A:Accession: H97570
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-315 <R>
A:Cross-references: GB:AE007869; PIDN:AAK87521.1; PID:g15156852; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C 3215
A:Map position: circular chromosome

```

Query Match      21.4%; Score 273; DB 2; Length 315;
Best Local Similarity 26.0%; Pred. No. 4e-16;
Matches 75; Conservative 47; Mismatches 84; Indels 82; Gaps 4;

QY      9  INLVSGGAAGKTAHIGVLKAINELGIRVRALSGVSAGAIVSFVASGVSPEGMFSLLKR 68
Db      37  IALALGGGAARGWAHIGVIRALDEAGVKIGIMWIAAGTSIGALVGCYLAG-----K 85

QY      69  VNWLKLPKFKPKLK-----GLIGWEKAIKRFLEEVLPFRRIEKELEIPYICAT 115
Db      86  LDELESEPARSLTMRRIAGLLDLDTTGGGGLFGGKRLTKRMOEHLGRLVENLEHPIAVAT 145

QY     116  DLYSGRALYLSGEGSLIPALLGSCAIPGIFEPFVRYKYNLLVDGGIVNNLPV----- 165
Db     146  ELRTGHEVYHIHQGDVLVALRSYALPGIFEPVQCGNRTLLDGLALVNPVPVSVCRAYEQAL 205

QY     166  -----EPQSSGIPTVCVDVLPIEBEKOIKNLH 194
Db     206  VVAVNINLDFGRSAVFKHAASPOGGTTPPAETAPRGLPGVWVQAFNIQQR----- 258

QY     195  ILLRSGFJAVRNSNKRKEFCDLVIVPELEBEFTPLDVRKADQIMERGY 242
Db     259  -----LSRSRLAGDPDPLMLHPRINDIGLSGFHFRASRAIDRGY 296

```

RESULT 10
H69874
conserved hypothetical protein ylbK - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: H69874
C:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Bertsch
C.; Bron, S.; Brouillet, S.; Bruschii, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chao
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Brington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997

Db 43 VGLALGGGASKGFPAHGIITKVKENGIPKVVATGVSAGSIYGSJFASGMSFDRLEBAEI 102

Qy 69 VNNLKLFFKPPKPKLIGMEKARFLEBVLPEYRIEKEIFPVYCATDLYSGALVYSEG 128

Db 103 LCKTDLVDLTLTSSGIRKEKONTNKRVGROIQCPPIKFAAVXDPEFGKVAANOG 162

Qy 123 SLIPALGSCAIPGTEPEYKNTLLVDGIVNNLPVEPQESGIPYCDVLPDPEKD 188

Db 163 MGOQVAFASAIENVPQPIGRHYVVGSLQPVPSAAROGANFYIADVISARPKN 222

Qy 189 KNIILHLRSPF-----LAVRSNKKRKEF--CDLIVPEL 223

Db 223 IS-----QGFFSYIDQTLNMTSVSALONEGADVIXQV 258

RESULT 12

F82852

conserved hypothetical protein XF0066 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: F82852

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequences

Nature 406, 151-157, 2000

A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A>Note: for a complete list of authors see reference number A59328 below

A:Accession: F82852

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-355 <SIM>

A:Cross-references: GB:AE003860; GB:AE003849; NID:99104830; PIDN:AF82879.1; GSPDB:GN001

A:Experimental source: strain 9a5c

R:Simpon, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvares, R.; A. Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H. as-Neto, R.; Docena, C.; El-Porri, H.; Facincani, A.F.; Ferreira, A.J.S.

submitted to Genbank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Klajima, J.P.; Kleger, J.E.; Kurame, E.B.; Laish chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E. A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A. Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Savasch A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tsunako, M.H.; Vailada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF0066

Query Match 21.1%; Score 268.5; DB 2; Length 395;

Best Local Similarity 29.4%; Pred. No. 1.3e-15;

Matches 74; Conservative 48; Mismatches 113; Indels 17; Gaps 4;

Db 9 INDVLSGAAKGIHAGVILKAINELGIRVALSGVSAIYVFAAGSYSEGFSLIKR 68

Qy 96 IGLALGGGAAKGFPAHGIITKVKENGIPKVVATGVSAGSIYGSJFASGMSFDRLEBAEI 152

Db 69 -----VNNLKLFFKPPKPKLIGMEKARFLEBVLPEYRIEKEIFPVYCATDLYSG 120

Qy 153 AVAMDESINRDLRFPS-----GLVQGGKTLQNYNELVGGKPIREKLNPFPAITRLDGD 207

Db 121 RAIVLSGSGIIPALGSCAIPGTEPEYKNTLLVDGIVNNLPVEPQESGIPYCDVLPDPEKD 179

Qy 208 QRTMFRKGVGAVASASCSIPGVFPVITGATHYDGSITSPVVDARQGLADPVAIVD 267

Db 180 VLPIEPEKQIKNLIHLRSPFLAVRSNKKRKEFCDLIVPELEFTPLDVRRAQIME 239

Qy 268 ISKANGKRPDGGIVGVNQSISIMQHGEALKRANIVHPKVLIDGSTDFDORNAITL 327

Db 240 RGIVKALEVISE 251

Qy 328 EGERATVIVIRE 339

RESULT 13

G85703

hypothetical protein ychK [imported] - Escherichia coli (strain O157:H7, substrain E

C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: G85703

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.T.; Ma iller, L.; Grobeck, B.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apod Nature 409, 529-533, 2001

A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85460; MUID:21074935; PMID:111206531

A:Accession: G85703

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-314 <STG>

A:Cross-references: GB:AE005174; NID:912514952; PIDN:AAG56091.1; GSPDB:GN00145; UWGP A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: ychK

Query Match 20.3%; Score 258.5; DB 2; Length 314;

Best Local Similarity 25.3%; Pred. No. 7.3e-15;

Matches 78; Conservative 49; Mismatches 106; Indels 75; Gaps 8;

Db 3 LRKPEINLVSGGAAGIAGITGILKAINELGIRVALSGVSAIYVFAAGSYSEGFSLIKR 62

Qy 14 MKKI-KIGALGSGAARMSHIGVINALKKGVEIDIVAGCSIGSLVGAAYAC-----DR 67

Db 63 FSLIKRVNMLKFFKPPKPKL-----GLIGMEKARFLEBVLPEYRIEKEIFPVYCATDLYSG 113

Qy 68 ISALE--DWVTSFSGVDTLRIMDSWQGGILNGERFVQYRIMETBENCSRFAAV 125

Db 114 ATDLYSGALYSGSLIPALGSCAIPGTEPEYKNTLLVDGIVNNLPVEPQESGIPYCDVLPDPEKD 173

Qy 126 ATNLSGTREIMFTGDLHLAIRASCSTIPGLMAPVANHGWLVGAVVNPPISTLRALGA 185

Db 174 PTVCDVLPPIPEKQIKNLIHL-----LRSPF-----LAVRSNKKRKEF--CDLIVPELEFTPLDVRRAQIME 243

Qy 166 DIVIN-----VDLQDHAIHQDLDLSTFVNSBENSNDISPLWIAKLERGSTTTR 237

Db 201 -----FLAVRSNKKRKEF--CDLIVPELEFTPLDVRRAQIME 243

Qy 238 AVTAPLATEIMTTSIOVLNKLKRNMAVGPDLIOPVQPISTLDFRHAHAIAAGOL 297

Db 244 KALEVISE 251

Qy 298 AVEKGMDE 305

RESULT 14

H90845

hypothetical protein ECs1736 [imported] - Escherichia coli (strain O157:H7, substrain E

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: H90845

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, (gaawara, N.; Yasunaga, T.; Kunita, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001

A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and A:Reference number: A99629; MUID:21156233; PMID:11258796

A:Accession: H90845

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-314 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA835159.1; PID:G13361201; GSPDB:GN00154 A:Experimental source: strain O157:H7, substrain RMD 0509952

C:Genetics:

A:Gene: ECs1736

Query Match 20.3%; Score 258.5; DB 2; Length 314;

Best Local Similarity 25.3%; Pred. No. 7.3e-15;

Search completed: June 19, 2003, 15:04:47
Job time : 29 secs

```
Matches 78; Conservative 49; Mismatches 106; Indels 75; Gaps 8;
Qy 3 LRKEEINVLGSGAAGIAHIGVLKAINELGIRVALSGVSAGAIVSFVAGSYSGM 62
Db 14 MRKI-KIGLALGSGAARGNSHIGVINALKKGIEIDIVAGCSIGSLVGAAYAC-----DR 67
Qy 63 FSLKRVNMLKFKFKPLK-----GLIGWEKAIKFLFLEVLPRRIEKLPIPTYIC 113
Db 68 LSALE--DWTFSYMDVLRMLDLSWQGGLLRGFRVNOYREIMPETEINCSRRFAAV 125
Qy 114 ATDLYSGHALYLSGSLIPALGSCAIPGIFEPVEYKYNLLVDGGIVNNLFPVEFQESGI 173
Db 126 ATNLSTGRELWFTGDLHLAIRASCIPGLMAPVAHNGYWLVDGAVNVPVPSLTRAIGA 185
Qy 174 PTVCDVLPTEPEKDKNHLIL--LRSF----- 200
Db 186 DIVIA-----VDLQDAHLMOQDLSLNFVNSENSGDSLPWHARKLERLGSITRR 237
Qy 201 -----FLAVRSNSEKRKEFC-----DLVIVPELEEFPTLDVRKADQIMERGYI 243
Db 238 AVTAPTATEIMTTSIQVLENRLKRNRMAGDPDILIQVPCQISTLDFHRAHAIAAAGQL 297
Qy 244 KALEVLESE 251
Db 298 AVEKKMDE 305

RESULT 15
AH0649
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C:Accession: AH0649
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
  T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
  S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
  et al.
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar
  Typhi
A:Reference number: AP0502; PMID:11677608
A:Accession: AH0649
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-301 <PA3>
A:Cross-references: GH:AL513382; PIDN:CAD08379.1; PID:gl6502423; GSPDB:GN00176
C:Genetics:
A:Gene: STY1296

Query Match 20.2%; Score 257.5; DB 2; Length 301;
Best Local Similarity 26.3%; Pred. No. 8.4e-15;
Matches 79; Conservative 49; Mismatches 113; Indels 59; Gaps 7;
Qy 3 LRKEEINVLGSGAAGIAHIGVLKAINELGIRVALSGVSAGAIVSFVAGSYSGM 62
Db 1 MRKM-KIGLALGSGAARGNSHIGVINALKKGIEIDIVAGCSIGSLVGAAYAC-----NK 54
Qy 63 FSLKRVNMLKFKFKPLK-----GLIGWEKAIKFLFLEVLPRRIEKLPIPTYIC 113
Db 55 LSALEQ--WVCFSYMDVLRMLDLSWQGGLLRGFRVNOYREIMPETEINCSRRFAAV 112
Qy 114 ATDLYSGHALYLSGSLIPALGSCAIPGIFEPVEYKYNLLVDGGIVNNLFPVEFQESGI 173
Db 113 ATNLSTGRELWFTGDLHLAIRASCIPGLMAPVAHNGYWLVDGAVNVPVPSLTRAIGA 172
Qy 174 PTV-----CVDVLPTEPEKDKNHLILRSFP----- 201
Db 173 DIVIAVDLQDAHLMOQDLSLNFVNSENSGDSLPWHARKLERLGSITRR 232
Qy 202 -----LAVRSNSEKRKEFC-----DLVIVPELEEFPTLDVRKADQIMERGYIKALEVLESE 251
Db 233 EIMTTSIQVLENRLKRNRMAGDPDILIQVPCQISTLDFHRAHAIAAAGQLAVEKKMDE 292
```


DR WPI: 1997-425035/39.
 DR N-PSDB; AAT79324.
 XX Nucleic acid encoding heat stable esterase from thermophilic
 PT bacteria - which is active in organic solvents, useful in cheese or
 PT paper manufacture, and to study plant resistance to disease
 XX
 XX Claim 19; Page 56-57; 113pp; English.
 XX
 CC This protein comprises the thermostable esterase 28LC of Aquifex
 CC pyrophilus, a Gram-negative, strictly chemolithoautotrophic knall
 CC gas marine bacterium which grows optimally at 85 deg C and pH
 CC 6.8. Newly identified polynucleotides (AAT79321-30) encoding claimed
 CC esterases (AAW23069-77, AAW23088) were recovered from genomic gene
 CC libraries. They can be used for recombinant production of the
 CC enzymes in host cells, and as probes to identify related sequences.
 CC The esterases are stable at high temperature and in organic
 CC solvents, making them superior for use in production of pure chiral
 CC compounds used in pharmaceutical, agricultural and other chemical
 CC industries. A method is claimed for transferring an amino group
 CC from an amino acid to an alpha-keto acid using a claimed esterase.
 CC The enzymes may also be useful as ripening starters in cheese making,
 CC in lignin removal in paper and pulp manufacture, in carbohydrate
 CC derivative synthesis, in fermentable sugar production from
 CC lignocellulosic waste, in the study of plant wall structure, plant
 CC resistance to disease and organic matter decomposition and to select
 CC plants bred for production of highly degradable animal feeds.
 XX
 SQ Sequence 251 AA;

Query Match 100.0%; Score 1273; DB 18; Length 251;
 Best Local Similarity 100.0%; Pred. No. 9.8e-133;
 Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LRLKFEINLVLSGGAAGIAHIGVLKAINELGIRVRLSGVSAIAVSFYASGYSP 60
 DB 1 LRLKFEINLVLSGGAAGIAHIGVLKAINELGIRVRLSGVSAIAVSFYASGYSP 60
 QY 61 GMFSILKRVNWLKLPKPPKGLIGWEKATRFLEVLVPRRIEKLPTTVCATDLYSG 120
 DB 61 GMFSILKRVNWLKLPKPPKGLIGWEKATRFLEVLVPRRIEKLPTTVCATDLYSG 120
 QY 121 RALYLSGSLIPALLGSCAIPGIFEPVEYKNVLLVDGGIVNNLPVEPFOESGIPVCVDV 180
 DB 121 RALYLSGSLIPALLGSCAIPGIFEPVEYKNVLLVDGGIVNNLPVEPFOESGIPVCVDV 180
 QY 181 LPIEPEKDKNHLHLLRSFPLAVRSNKRKFCDLVIVPELEBPTPLDVRKADQIMR 240
 DB 181 LPIEPEKDKNHLHLLRSFPLAVRSNKRKFCDLVIVPELEBPTPLDVRKADQIMR 240
 QY 241 GYIKALEVLS 251
 DB 241 GYIKALEVLS 251

RESULT 2
 AAW23074
 ID AAW23074 standard; Protein: 249 AA.
 XX
 XX AAW23074;
 XX
 DT 17-FEB-1998 (first entry)
 XX
 DE Aquifex esterase VF5-34LC.
 XX
 KW Esterase; thermostable enzyme; ester; chiral compound; cheese;
 KW pulp; paper; lignin removal; sugar; lignocellulose;
 KW disease resistance; feedstuff.
 XX
 OS Aquifex sp. strain VF5.
 XX
 FT Key Location/Qualifiers
 Misc-difference 99

PT XX /note= "encoded by TAA"
 FN WO9730160-A1.
 XX 21-AUG-1997.
 PD 11-FEB-1997; 97WO-US02039.
 XX 16-FEB-1996; 96US-0602359.
 PR (RECO-) RECOMBINANT BIOCATALYSIS INC.
 FA Callen W, Kosmotka A, Link S, Maffia AM, Murphy D;
 PI Reid J, Robertson DE, Swanson RV, Warren PV;
 XX WPI: 1997-425035/39.
 DR N-PSDB; AAT79327.
 XX Nucleic acid encoding heat stable esterase from thermophilic
 PT bacteria - which is active in organic solvents, useful in cheese or
 PT paper manufacture, and to study plant resistance to disease
 XX
 XX Claim 19; Page 59-60; 113pp; English.
 CC This protein comprises the thermostable esterase VF5-23LC of
 CC Aquifex VF5, a marine strictly chemolithoautotrophic knall
 CC gas bacterium that grows optimally at 85 deg C and pH 6.8.
 CC Newly identified polynucleotides (AAT79321-30) encoding claimed
 CC esterases (AAW23069-77, AAW23088) were recovered from genomic libraries.
 CC They can be used for recombinant production of the enzymes in host
 CC cells, and as probes to identify related sequences. The esterases
 CC are stable at high temperature and in organic solvents, making them
 CC superior for use in production of pure chiral compounds used in
 CC pharmaceutical, agricultural and other chemical industries. A
 CC method is claimed for transferring an amino group from an amino
 CC acid to an alpha-keto acid using a claimed esterase. The enzymes
 CC may also be useful as ripening starters in cheese making, in lignin
 CC removal in paper and pulp manufacture, in carbohydrate derivative
 CC synthesis, in fermentable sugar production from lignocellulosic
 CC waste, in the study of plant wall structure, plant resistance to
 CC disease and organic matter decomposition and to select plants bred
 CC for production of highly degradable animal feeds.
 XX
 SQ Sequence 249 AA;
 Query Match 71.8%; Score 913.5; DB 18; Length 249;
 Best Local Similarity 72.0%; Pred. No. 8.5e-93;
 Matches 177; Conservative 31; Mismatches 37; Indels 1; Gaps 1;
 QY 1 LRLKFEINLVLSGGAAGIAHIGVLKAINELGIRVRLSGVSAIAVSFYASGYSP 60
 DB 5 LKLRFEVNLVLSGGAAGIAHIGVLKALEELGKVKRLSGVSAIAVSFYASGYTPD 64
 QY 61 GMFSILKRVNWLKLPKPPKGLIGWEKATRFLEVLVPRRIEKLPTTVCATDLYSG 120
 DB 65 EMLKLLKEVNLKLPKFPKPPKGLMGWEKAEFLKELGVKRLLEDNLTPTTLCADLYTG 124
 QY 121 RALYLSGSLIPALLGSCAIPGIFEPVEYKNVLLVDGGIVNNLPVEPFOESGIPVCVDV 180
 DB 125 KALYFCRGDLIPVLLGSCSIPGIFEPVEYENFLVDGGIVNNLPVEPLEKPKPIIGVDV 184
 QY 181 LPIEPEKDKNHLHLLRSFPLAVRSNKRKFCDLVIVPELEBPTPLDVRKADQIMR 240
 DB 185 LPITQERKKNLHLLRSFPLAVRSNKRKFCNNVIRPPLPESFPLDVRKADQIMR 243
 QY 241 GYIKAL 246
 DB 244 GDMRAL 249

RESULT 3
 AAY38779
 ID AAY38779 standard; Protein: 300 AA.

```
XX AC AAY38779;
XX DT 08-OCT-1999 (first entry)
XX DE Neisseria meningitidis strain A antigen encoded by ORF137.
XX KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX KW treatment; Neisseria infection; meningitis; septicemia; gonorrhea.
XX OS Neisseria meningitidis.
XX PN WO924578-A2.
XX PD 20-MAY-1999.
XX PF 09-OCT-1998; 98WO-IB01665.
XX PR 01-SEP-1998; 98GB-0019016.
XX PR 06-NOV-1997; 97GB-0023516.
XX PR 14-NOV-1997; 97GB-0024190.
XX PR 18-NOV-1997; 97GB-0024386.
XX PR 27-NOV-1997; 97GB-0025158.
XX PR 10-DEC-1997; 97GB-0026147.
XX PR 14-JAN-1998; 98GB-0000759.
XX PA (CHIR-) CHIRON SPA.
XX PI Grandt G, Masignani V, Pizzo M, Rappuoli R, Scarlato V;
XX DR WPI; 1999-327407/27.
XX DR N-PSDB; AA212214.
XX PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX PS diagnosis, treatment and prevention of infection
XX PS Claim 4; Page 324; 524pp; English.
XX CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
XX CC and N. gonorrhoeae antigenic proteins. They are encoded by open
XX CC reading frames (ORFs) AK211972-212358. The antigenic proteins,
XX CC their fragments, their nucleic acids and antibodies are used for
XX CC diagnosis, prevention (as vaccines) or treatment of Neisseria
XX CC infections, such as meningitis, septicemia and gonorrhea. Both
XX CC organisms are closely related. Fragments of the nucleic acids
XX CC are useful as hybridisation probes and antisense reagents.
XX SQ Sequence 300 AA;
XX
Query Match 21.2%; Score 269.5; DB 20; Length 300;
Best Local Similarity 31.4%; Pred. No. 3; 9e-21;
Matches 70; Conservative 40; Mismatches 98; Indels 15; Gaps 3;
QY 9 INVLVSGGAKGIAGIIVGLKALNEIGIRVRALSGVSAQIVSVFYASGSPGCMFSLKR 68
DB 46 VGLALGGAGSKGFARHGIIRKLVKENGIPKVVITGSGISVGSFASGMPDLBLEAEI 105
QY 69 VNMILKFKKPPKGLIGMEKALRFLBEVLPRYIEKLEIPTYICANDLYSGALYLSG 128
DB 106 LKGTDLVDLTLSGPIKGEKQNYINRKVGRRIOQPIKFAAVADDFETGAAVAFNG 165
QY 129 SLIPALGSCAIPGIEPEVEYKAYLLVDGIVNNLPPEPQESGIPYVCVDFPIPERD 188
DB 166 NAGQAVRASAALPNVFPVITIGRHTYVDGILSQPVFASARRXXXXXVIAVDISARPSGN 225
QY 189 IKNIILILSRF-----LAVRSNSEKKEF--CDLYIVPEL 223
DB 226 IS-----QGFPSTYDQTLNWSVSAQNELGQADVIRKPOV 261
```

RESULT 4
AAY38778
ID AAY38778 standard; Protein; 300 AA.

```
XX AC AAY38778;
XX DT 08-OCT-1999 (first entry)
XX DE Neisseria meningitidis antigen encoded by ORF137.
XX KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX KW treatment; Neisseria infection; meningitis; septicemia; gonorrhea.
XX OS Neisseria meningitidis.
XX PN WO924578-A2.
XX PD 20-MAY-1999.
XX PF 09-OCT-1998; 98WO-IB01665.
XX PR 01-SEP-1998; 98GB-0019016.
XX PR 06-NOV-1997; 97GB-0023516.
XX PR 14-NOV-1997; 97GB-0024190.
XX PR 18-NOV-1997; 97GB-0024386.
XX PR 27-NOV-1997; 97GB-0025158.
XX PR 10-DEC-1997; 97GB-0026147.
XX PR 14-JAN-1998; 98GB-0000759.
XX PA (CHIR-) CHIRON SPA.
XX PI Grandt G, Masignani V, Pizzo M, Rappuoli R, Scarlato V;
XX DR WPI; 1999-327407/27.
XX DR N-PSDB; AA212213.
XX PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX PS diagnosis, treatment and prevention of infection
XX PS Claim 4; Page 323; 524pp; English.
XX CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
XX CC and N. gonorrhoeae antigenic proteins. They are encoded by open
XX CC reading frames (ORFs) AK211972-212358. The antigenic proteins,
XX CC their fragments, their nucleic acids and antibodies are used for
XX CC diagnosis, prevention (as vaccines) or treatment of Neisseria
XX CC infections, such as meningitis, septicemia and gonorrhea. Both
XX CC organisms are closely related. Fragments of the nucleic acids
XX CC are useful as hybridisation probes and antisense reagents.
XX SQ Sequence 300 AA;
XX
Query Match 21.1%; Score 268.5; DB 20; Length 300;
Best Local Similarity 31.4%; Pred. No. 5; 1e-21;
Matches 70; Conservative 41; Mismatches 97; Indels 15; Gaps 3;
QY 9 INVLVSGGAKGIAGIIVGLKALNEIGIRVRALSGVSAQIVSVFYASGSPGCMFSLKR 68
DB 46 VGLALGGAGSKGFARHGIIRKLVKENGIPKVVITGSGISVGSFASGMPDLBLEAEI 105
QY 69 VNMILKFKKPPKGLIGMEKALRFLBEVLPRYIEKLEIPTYICANDLYSGALYLSG 128
DB 106 LKGTDLVDLTLSGPIKGEKQNYINRKVGRRIOQPIKFAAVADDFETGAAVAFNG 165
QY 129 SLIPALGSCAIPGIEPEVEYKAYLLVDGIVNNLPPEPQESGIPYVCVDFPIPERD 188
DB 166 NAGQAVRASAALPNVFPVITIGRHTYVDGILSQPVFASARRQGANRVIAVDISARPSGN 225
QY 189 IKNIILILSRF-----LAVRSNSEKKEF--CDLYIVPEL 223
DB 226 IS-----QGFPSTYDQTLNWSVSAQNELGQADVIRKPOV 261
```

RESULT 5
AAY38780
ID AAY38780 standard; Protein; 300 AA.

```
XX AC AAY38780;
XX DT 08-OCT-1999 (first entry)
XX DE Neisseria gonorrhoeae antigenic protein encoded by ORF137.
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX OS Neisseria gonorrhoeae.
XX PN W09924578-A2.
XX PD 20-MAY-1999.
XX PF 09-OCT-1998; 98WO-IB01665.
XX PR 01-SEP-1998; 98GB-0019016.
XX PR 06-NOV-1997; 97GB-0023516.
XX PR 14-NOV-1997; 97GB-0024190.
XX PR 18-NOV-1997; 97GB-0024386.
XX PR 27-NOV-1997; 97GB-0025158.
XX PR 10-DEC-1997; 97GB-0026147.
XX PR 14-JAN-1998; 98GB-0000759.
XX PA (CHIR-) CHIRON SPA.
XX PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX WPI; 1999-327407/27.
XX DR N-PSDB; AA212215.
XX PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX diagnosis, treatment and prevention of infection
XX PS Claim 4; Page 325; 524pp; English.
XX CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC and N. gonorrhoeae antigenic proteins. They are encoded by open
CC reading frames (ORFs) AA211972-Z12358. The antigenic proteins,
CC their fragments, their nucleic acids and antibodies are used for
CC diagnosis, prevention (as vaccines) or treatment of Neisseria
CC infections, such as meningitis, septicaemia and gonorrhea. Both
CC organisms are closely related. Fragments of the nucleic acids
CC are useful as hybridisation probes and antisense reagents.
XX SQ Sequence 300 AA;
XX Query Match 20.4%; Score 259.5; DB 20; Length 300;
XX Best Local Similarity 31.4%; Pred. No. 5.1e-20;
XX Matches 70; Conservative 40; Mismatches 98; Indels 15; Gaps 3;
QY 9 INVLGCAAKGIAHIGVLKAINELGIRVRLSGVSAIAIVSYFASGYSPGWFSLKR 68
DB 46 VALALGGGASKGFAHIGIVKVKENGIPKVVVTGTSAGSIVGSLLAGSPDRLELEAEI 105
QY 69 VNMVLEKFKPLKGLICWKAIRFLERVLPRYRIEKLPTTYICATDLYSGRALVSEK 128
DB 106 LGKTDVLDTLSTSGFIKGLQYINKKVGRIQQFPKFAVAIDFTGKAVAFNQ 165
QY 129 SLIPALLGSCAIPGIFEPVYKYNLLVDGGIVNNLVPFQESGIPVVDVLPFPEKD 188
DB 166 NAGQAVRASAAIPNVPFVIGRHKYVDGGLSQPVVSAARRQGANFVIAVDISARFSKN 225
QY 189 IKNILHILLRSPF-----LAVRSNSEKKEF--CDLVIVPEL 223
DB 226 VG-----QGFFSYLDQTLNMSVSVLQNELGQADVIVKPVQV 261
RESULT 6
AAY70474
ID AAY70474 standard; Protein; 1327 AA.
```

```
XX AC AAY70474;
XX DT 04-JUL-2000 (first entry)
XX DE Human cyclic nucleotide-associated protein-2 (CNAP-2).
XX KW Cyclic nucleotide-associated protein-2; CNAP-2; human; cytostatic;
KW anti-arteriosclerotic; hepatotropic; anti-leukaemic; anti-inflammatory;
KW immunomodulatory; anti-asthmatic; anti-anemic; anti-diabetic; diagnosis;
KW anti-sclerotic; dermatological; neuroprotective; anti-epileptic; cancer;
KW anti-Alzheimer's; anti-Parkinsonian; cerebroprotective; ophthalmological;
KW anti-infertility; anti-allergic; vasotropic; immunosuppressive;
KW hypotensive; gene therapy; prevention; treatment; arteriosclerosis;
KW cell proliferative disorder; autoimmune/inflammatory; diabetes mellitus;
KW neurological; vision; reproductive; smooth muscle.
XX OS Homo sapiens.
XX Key Location/Qualifiers
XX Peptide 1..34
XX Protein 35..1327
XX /label= Mature CNAP-2
XX /note= "Shares 24% identity to Aquifex pyrophilus
XX esterase 28LC"
XX Modified-site 68
XX /note= "Potential phosphorylation site"
XX Modified-site 1225
XX /note= "Potential phosphorylation site"
XX Modified-site 73
XX /note= "Potential phosphorylation site"
XX Modified-site 125
XX /note= "Potential phosphorylation site"
XX Modified-site 220
XX /note= "Potential phosphorylation site"
XX Modified-site 326
XX /note= "Potential phosphorylation site"
XX Modified-site 357
XX /note= "Potential phosphorylation site"
XX Modified-site 386
XX /note= "Potential phosphorylation site"
XX Modified-site 400
XX /note= "Potential phosphorylation site"
XX Modified-site 432
XX /note= "Potential phosphorylation site"
XX Modified-site 455
XX /note= "Potential phosphorylation site"
XX Modified-site 560
XX /note= "Potential phosphorylation site"
XX Modified-site 600
XX /note= "Potential phosphorylation site"
XX Modified-site 780
XX /note= "Potential phosphorylation site"
XX Modified-site 784
XX /note= "Potential phosphorylation site"
XX Modified-site 997
XX /note= "Potential phosphorylation site"
XX Modified-site 1113
XX /note= "Potential phosphorylation site"
XX Modified-site 1121
XX /note= "Potential phosphorylation site"
XX Modified-site 1171
XX /note= "Potential phosphorylation site"
XX Modified-site 1251
XX /note= "Potential phosphorylation site"
XX Modified-site 1274
XX /note= "Potential phosphorylation site"
XX Modified-site 1285
XX /note= "Potential phosphorylation site"
XX Modified-site 1299
XX /note= "Potential phosphorylation site"
XX Modified-site 1301
XX /note= "Potential phosphorylation site"
```

FT	Modified-site	/note= "Potential phosphorylation site"
FT	1323	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	92	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	236	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	319	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	547	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	634	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	699	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	816	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	894	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	910	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	1220	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	1230	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	392	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "Potential phosphorylation site"
FT	1019	/note= "Potential phosphorylation site"
FT	Modified-site	/note= "N-glycosylated"
FT	1040	/note= "N-glycosylated"
FT	Modified-site	/note= "N-glycosylated"
FT	1228	/note= "N-glycosylated"
FT	Binding-site	/label= cNMP-binding_domain
FT	144..269	/label= cNMP-binding_domain
FT	Binding-site	/label= cNMP-binding_domain
FT	573..696	/label= cNMP-binding_domain
FT	Domain	/label= Transmembrane_domain
FT	10..30	/label= Transmembrane_domain
FT	Region	/note= "Resembles cyclic-nucleotide binding domain proteins"
FT	605..628	/note= "Resembles cyclic-nucleotide binding domain proteins"
FT	Region	/note= "Resembles cyclic-nucleotide binding domain proteins"
FT	643..676	/note= "Resembles cyclic-nucleotide binding domain proteins"
FT	Region	/note= "Resembles cyclic-nucleotide binding domain proteins"
XX	WO200014248-A1.	
XX	16-MAR-2000.	
XX		
XX	03-SEP-1999;	99WO-US20287.
XX	04-SEP-1998;	98US-0148904.
XX	(INCY-) INCYTE PHARM INC.	
XX		
XX	Hillman JL, Yue H, Guegler KJ, Corley NC, Patterson C, Tang YT;	
XX	WPI, 2000-256994/22.	
XX	N-PDB; AA251683.	
XX		
XX	Isolated cyclic nucleotide associated proteins useful for preventing,	
XX	diagnosing and treating cell proliferative, autoimmune/inflammatory,	
XX	neurological, vision, reproductive and smooth muscle disorders -	
XX	Disclosure; Page 64-67; 78pp; English.	
XX		
XX	The present sequence is a human cyclic nucleotide-	
XX	associated protein-2 (CNP-2), identified in Incyte clone 3149674,	
XX	that is isolated from ADRENOMD4 cDNA library. It is expressed in	
XX	nervous, reproductive, cardiovascular and haematopoietic/immune tissues	
XX	CNP sequences may be used for prevention, treatment and diagnosis of	
XX	diseases associated with altered CNP expression such as, cell	

CC proliferative disorders (e.g. arteriosclerosis, cirrhosis, leukaemia,
CC lymphoma and cancer of the breast, prostate, lung and brain), autoimmune/
CC inflammatory disorders (e.g. asthma, anaemia, diabetes mellitus, multiple
CC sclerosis and psoriasis), neurological disorders (e.g. epilepsy,
CC Alzheimer's/Parkinson's disease and strokes), vision disorders (e.g.
CC conjunctivitis, glaucoma, cataracts and retinitis pigmentosa),
CC reproductive disorders (e.g. infertility, uterine fibroids, ectopic
CC pregnancies and impotence) and smooth muscle disorders (e.g. angina,
CC anaphylactic shock, Kearns-Sayre syndrome and hypertension). The
CC coding sequence can be used for gene therapy.

SQ Sequence 1327 AA;

Query Match 17.6%; Score 223.5; DB 21; Length 1327;
Best Local Similarity 26.5%; Pred. No. 4,5e-15;
Matches 74; Conservative 37; Mismatches 113; Indels 55; Gaps 5;

QY 9 INLVISGANGAIHIGVLKAINELGIRVPAISGVASAGAIIVSEFYASGY----- 58
| | | | | : : : | | | | | : : : | | |
Dd 931 IALVVGCGAGACSGSITGLKALEBAQVPDLVGSTISGFICALVAEBRSARTQRARE 990
| | | | | : : : | | | | | : : : | | |

QY 59 -PEGNPSLTKXYNNMLKFKEKPDKLGIEKKAIRPLEEVLTPRRERKLEPTTYICANDL 117
| | | | | : : : | | | | | : : : | | |
Dd 991 WAKSMSTVLEPV-----LDLTYPVTSMFTGSAPNRKSIHRVFODKOLEDMLPEFVVTTDI 1045
| | | | | : : : | | | | | : : : | | |

QY 118 YSGRALYLSESLIPALISCAIPGIPEPV--EKRYVLLVDGGIIVNNLPVBPDSGIPT 175
| | | | | : : : | | | | | : : : | | |
Dd 1046 TYSANRKHQDSLMRYVASWTLISGYPLPCDPCKDHLLMDGGYINNLPADIAFSMGANT 1105
| | | | | : : : | | | | | : : : | | |

QY 176 VCVDVLPPIEPERKDKNILHLHSFFLAVSNSEKKKECDLYIVELEE----- 225
| | | | | : : : | | | | | : : : | | |
Dd 1106 VALIDVQSDETDISTYGDLSGWWLLMKRLN----PMADKKYVPDMAEQSRLAYVSC 1160
| | | | | : : : | | | | | : : : | | |

QY 226 -----FTPLDVRKADIMERGY 242
| | | | | : : : | | | | | : : : | | |
Dd 1161 VRQLSEVVKSSGYCEYLREPDIQCFETMDFGKFDGIYDVGY 1199
| | | | | : : : | | | | | : : : | | |

RESULT 7
ABB58963
ID ABB58963 standard; Protein; 1389 AA.

AC ABB58963;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 3681.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical.
KM
OS Drosophila melanogaster.
XX
WO200171042-A2.
PN
PD 27-SEP-2001.
PF 23-MAR-2001; 2001WO-US09231.
PR 23-MAR-2000; 2000US-191637P.
PP 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW,
DR WPI; 2001-656860/75.
DR N-PDB; ABI03066.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -
XX

PS Disclosure; SEQ ID NO 3681; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from prokaryotes. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB161176-AB161175), expressed DNA sequences (AB161176-AB161175) and the encoded proteins (AB161176-AB161175).

CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1389 AA;

Query Match 17.1%; Score 217.5; DB 22; Length 1389;
Best Local Similarity 26.1%; Pred. No. 2.2e-14;
Matches 73; Conservative 44; Mismatches 108; Indels 55; Gaps 9;

QY 9 INLVSGGAAGIAHIGVLAINELGIRVRLSGVSAAGVSVFYASGYSPGCMFSL-- 66
DB 914 IGLVLGGGARGAAGHGMKAIQIAGIPVDMVGVSIGALMGALWCS----ERNITVTQ 969

QY 67 -----KRVNW-LKLFKRPPLKGLIGWEKAIKRLFEVLVPRRIEKLPTIYCATDLY 118
DB 970 KARENSKMTKWFLLQDLDTVITSMFGSGREFNKTIDTGDVSIEDLWIPYFTLTDTIT 1029

QY 119 SGRALYSEGLIIPALLGSCAIPGIFEPV--EYKNYLLVDGGIVNNLPVFPFQESGIP-T 175
DB 1030 ASCHRIHTNGSLWRYTRSSMSLSGYMPLCPDKGHLLDGGVYNNLPADVMMNLGAARI 1089

QY 176 VCVDV-----LP1EPEKDIKNILHLILRSFPLAV 204
DB 1090 IADVGSQDQDTLTNYGDDLSGWLKLYKKNPFTSPVKVPLDPIQSL-----AVSCV 1144

QY 205 RNSSE-KRKEFCDLVIVPELEFPTLDVRKADQIMERGI 243
DB 1145 RLEEVKNSDYCE-YIRPIDKYKTLAPGSPFDIRDVGIV 1183

RESULT 8
ID AAW23079 standard; Protein; 259 AA.
XX AAW23079;
AC AAW23079;
XX 17-FEB-1998 (first entry)
XX Whale mat sample 11.801 esterase es9.
XX Esterase; thermostable enzyme; ester; chiral compound; cheese;
XX pulp; paper; lignin removal; sugar; lignocellulose;
XX disease resistance; feedstuff.
XX Whale mat sample 11.801.
XX WO9730160-A1.
XX 21-AUG-1997.
XX 11-FEB-1997; 97WO-US02039.
XX 16-FEB-1996; 96US-0602359.
XX (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX Callen W, Kosmicka A, Link S, Maffia AM, Murphy D;
XX Reid J, Robertson DE, Swanson RV, Warren PV;
XX WPI; 1997-425035/39.
XX N-PSDB; AAT79332.
XX

PT Nucleic acid encoding heat stable esterase from thermophilic
PT bacteria - which is active in organic solvents, useful in cheese or
XX paper manufacture, and to study plant resistance to disease
PS Disclosure; Page 64-65; 113pp; English.

XX This protein comprises the whale mat sample 11.801 esterase es9.
CC Newly identified polynucleotides (AAT79321-40) encoding esterases
CC (AAW23069-88), some of which are claimed, can be used for
CC recombinant production of the enzymes in host cells. The esterases
CC are stable at high temperature and in organic solvents, making them
CC superior for use in production of pure chiral compounds used in
CC pharmaceutical, agricultural and other chemical industries. A
CC method is claimed for transferring an amino group from an amino acid
CC to an alpha-keto acid using a claimed esterase. The enzymes may
CC also be useful as ripening starters in cheese making, in lignin
CC removal in paper and pulp manufacture, in carbohydrate derivative
CC synthesis, in fermentable sugar production from lignocellulosic
CC waste, in the study of plant wall structure, plant resistance to
CC disease and organic matter decomposition and to select plants bred
CC for production of highly degradable animal feeds.

XX Sequence 259 AA;

Query Match 14.8%; Score 188.5; DB 18; Length 259;
Best Local Similarity 27.3%; Pred. No. 3.2e-12;
Matches 69; Conservative 43; Mismatches 118; Indels 23; Gaps 9;

QY 7 EETNLVSGGAAGIAHIGVLAINELGIRVRLSGVSAAGVSVFYASGYSPGCMFSL 66
DB 8 ENSLVLSGGGALGIAHIGVLAINELGIRVRLSGVSAAGVSVFYASGYSPGCMFSL 67

QY 67 KR-----VNNLKLKFKPPLKGLIGWEKAIKRLFEVLVPRRIEKLPTIYCATDLYSG-- 120
DB 68 KFSNVNENWIK-RSPSG--NSVVDNEKIAKIFDTLFKDKMTDTYIPLKLIATNLNGHK 124

QY 121 RALYLSSEGLI-PALLGSCAIPGIFEPVYKNVLLVDGGIVNNLPVFPFQESGIPVCVD 179
DB 125 KVFTASDDVLKDAIILSTWAIKGVFEHIIDGETTGDFLNCENLVNEATFNDV--LAVD 182

QY 180 VL-----PIEPEKDIK--NILHLILRSFPLAVRNSSE---KRKEFCDLVIVPELEFPT 228
DB 183 VMGENSEKAMPDNFFKTSNVLEMFESKMLFIYNTQTHIKNANKNIYLIBFTVKEYKT 242

QY 229 LDVRKADQIMERG 241
DB 243 FQPHKHKIRALG 255

RESULT 9
ID AAG22116 standard; Protein; 1402 AA.
XX AAG22116;
AC AAG22116;
XX 18-FEB-2002 (first entry)
XX Novel human diagnostic protein #22107.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX

PA (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX MPI: 2001-639362/73.
XX
XX N-PSDB: A8686303.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity
XX
XX Claim 20; SEQ ID No 52475; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridization probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. A860010-A8630377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pcc_sequences.
XX
XX Sequence 1402 AA;
XX
XX Query Match 14.6%; Score 186; DB 22; Length 1402;
XX Best Local Similarity 25.5%; Pred. No. 7.1e-11;
XX Matches 74; Conservative 42; Mismatches 116; Indels 58; Gaps 10;
XX
XX 3 LKKEBINLVSGG--AKGIAHGVKALINELGIRVALSGVAGAVSVFYASGYS-- 58
XX DB 965 LTRGNTALVLEGRGARGGSHIGVLALEBAGVPDLVGSTIGSFIALYAEBSAS 1024
XX QY 59 -----PEGMPSLKRVMWLKLFKKPPLKGLIGMEKARLEEVLP-YRIEQL 108
XX DB 1025 RTKQARAEKAKMTSVLEPV-----LDLTDAITSFTGSAANRSIHRYFPKSRFEDEL 1079
XX QY 109 PTY-ICATDLYSGRALYLSGSL--IPALLGSCA--IPGFEPEVRYNYLLVDGIVNN 162
XX DB 1080 PYFKKXPTDITASARGKDGSLMKRYVASMDLTSGVLPPLCDPKD--GHILMDGXYNN 1137
XX QY 163 LVEHPQESGIPITVGVVLPPIEPEDIKNIHLILRSFILAVERNS----- 208
XX DB 1138 LPADLARSMAKTAIVADGSDDETDLSTYGDSLFGMWLLMKRLNPMGLTRXRPDMAET 1197
XX QY 209 -----EKRRKPCDVLIVPELEEFPTLDRKADIMERGY 242
XX DB 1198 QPFGPNVSCVRLQEVVSSSYCE-YLAPPIDCFKTMDFGFDQIYDVGY 1246
XX
XX RESULT 10
XX ABP28552
XX ID ABP28552 standard; Protein; 283 AA.
XX
XX AC ABP28552;
XX
XX DT 02-JUL-2002 (first entry)
XX
XX DB Streptococcus polypeptide SEQ ID NO 6280.

KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
XX Streptococcus agalactiae.
XX
XX WO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB04789.
XX
XX 27-OCT-2000; 2000GB-0026333.
XX PR 24-NOV-2000; 2000GB-0028727.
XX PR 07-MAR-2001; 2001GB-0005640.
XX
XX PA (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX
XX Telford J, Naisnani V, Margarit Ros YI, Grandi G, Frazer C,
XX PI Tetteilin H;
XX
XX MPI: 2002-352536/38.
XX DB N-PSDB: A869183.
XX
XX Claim 1; Page 3791; 4525bp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GBS
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX CC the specification. The proteins have antibacterial and antiinflammatory
XX CC activity. (I), nucleic acids encoding (I), A866044-ABN71526 and
XX CC antibodies that bind (I) are used in the manufacture of medicaments for
XX CC the treatment or prevention of infection or disease caused by
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a
XX CC biological sample. (I) is used to determine whether a compound binds to
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX CC used as a vaccine or diagnostic composition. The disease caused by
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be
XX CC used in gene therapy. Antibodies to (I) are used for affinity
XX CC chromatography, immunoassays, and distinguishing/identifying
XX CC Streptococcus proteins.
XX
XX Sequence 283 AA;
XX
XX Query Match 13.2%; Score 168; DB 21; Length 283;
XX Best Local Similarity 23.7%; Pred. No. 6.8e-10;
XX Matches 66; Conservative 52; Mismatches 118; Indels 42; Gaps 8;
XX
XX 9 INLVSGGAAGIAHGVKALINELGIRVALSGVAGAVSVFYASGYSPEGMPSLKR 68
XX DB 6 VGLVLEGGKGLTAGLDAFLDAGIKIDGIVSVSGALFGVNFVRSRERL-----R 60
XX QY 69 VMMLKLFKKPPLKGLIGMEKARLEEVLPYRII-----EKLEIFTYICATD 116
XX DB 61 YN-KCIVSHPKVMSLSRSMFRFTGPFVNDFTYIVPMKLVDPDEAKKSSIDIFYVATE 118
XX QY 117 LYSGRALYLSGSL--IPALLGSCAIPGFEPEVRYNYLLVDGIVNNLPVEPQESGI 173
XX DB 119 MTSGAPREYFKIDSVFBOQETILRASALPVVSKWMDQCKYLDGSLSDSIPVFAKGLF 178
XX QY 174 -PTVCADVLPPIEPEDIKNI--ILHLRSFILAVERNSERKPCD-LIVPELEEFPTL 229
XX DB 179 DKLIVVMTPLNPKQKSSGRLYKTVIRKYPNPKATSNRYQGYNNLSLEKVMLEKTKGL 238
XX QY 230 -----DVRKADIMERGYIKALEVISE 251

Db 239 FAIRPSKSLVIGRLKPNPKLDSIYOLGMDKAKSVMP 276

RESULT 11

ABP29892 ID ABP29892 standard; Protein; 283 AA.

XX AC ABP29892;

XX DT 02-JUL-2002 (first entry)

XX DE Streptococcus polypeptide SEQ ID NO 8960.

XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;

XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.

XX OS Streptococcus agalactiae.

XX PN WO200234771-A2.

XX PD 02-MAY-2002.

XX PF 29-OCT-2001; 2001WO-GB04789.

XX PR 27-OCT-2000; 2000GB-0026333.

XX PR 24-NOV-2000; 2000GB-0028727.

XX PR 07-MAR-2001; 2001GB-0005640.

XX XX

XX PA (CHIR-) CHIRON SPA.

XX PA (GENO-) INST GENOMIC RES.

XX PI Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;

XX PI Tettelin H;

XX XX

XX DR WPI: 2002-352536/38.

XX DR N-PSDB; ABN70523.

XX XX

XX PT New Streptococcus protein for the treatment or prevention of infection

XX PT or disease caused by Streptococcus bacteria, such as meningitis, and

XX PT for detecting a compound that binds to the protein -

XX PS Claim 1; Page 4017; 4525pp; English.

XX CC The invention relates to a protein (ABP25413-ABP30895) from group B

XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in

XX CC the specification. The proteins have antibacterial and antiinflammatory

XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and

XX CC antibodies that bind (I) are used in the manufacture of medicaments for

XX CC the treatment or prevention of infection or disease caused by

XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a

XX CC biological sample. (I) is used to determine whether a compound binds to

XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be

XX CC used as a vaccine or diagnostic composition. The disease caused by

XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic

XX CC acid encoding (I) may be used to recombinantly produce (I) and may be

XX CC used in gene therapy. Antibodies to (I) are used for affinity

XX CC chromatography, immunoassays, and distinguishing/identifying

XX CC Streptococcus proteins.

XX SQ Sequence 283 AA;

Query Match 13.2%; Score 168; DB 23; Length 283;

Best Local Similarity 23.7%; Pred. No. 6.8e-10;

Matches 66; Conservative 52; Mismatches 118; Indels 42; Gaps 8;

QY 9 INLVLSGAAGKGIHGVLAINEIGIRVRALSGVAGAVSVFVAGSVSPGGMFSLIKR 68

Db 6 VGLVLEGGMEGLYTAGVLDALFDAGIKINDIVSVAGALFGVNFVSQRERAL-----R 60

QY 69 VNWKLKFKPKPLKGLIGWEKAIKPLEVLPIYRI-----EKLSPITYICATD 116

Db 61 YN--KKYLHPKYMSLRSWFTGCFNFKDFTYYEVPKLDVDFDDEAFKKSIDFYVATE 118

QY 117 LYSGRALYLSGSL----IPALLGSCAICFIEPEYKRYKLLVDGGIYNNLPVPPFQBSGI 173

Db 119 MTSKPEYFKIDSVPFQMEILRASALPWSKMDVQCKYLDGGLSDSIEVPDFARLGF 178

QY 174 -PTVCVDVLPFIEPEKIDKN--ILHILLRSFFLAVRSNKKKFCFCD-LVIVPELEETPL 229

Db 179 DKLIVVMTRPLNYQKKPSSGRLYKTKRYPNFVKTSNRYYQYNNLSLEKMSLEKTGDL 238

QY 230 -----DVRKADQIMERGVIKALEVLSE 251

Db 239 FAIRPSKSLVIGRLKPNPKLDSIYOLGMDKAKSVMP 276

RESULT 12

ABB52902 ID ABB52902 standard; Protein; 356 AA.

XX AC ABB52902;

XX DT 11-FEB-2002 (first entry)

XX DE Escherichia coli polypeptide SEQ ID NO 1215.

XX KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;

XX KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;

XX KW systemic infection; non-diarrhoeal infection; septicemia;

XX KW pyelonephritis; antibiotic resistance.

XX OS Escherichia coli.

XX PN WO200166572-A2.

XX PD 13-SEP-2001.

XX PF 12-MAR-2001; 2001WO-EP03445.

XX PR 10-MAR-2000; 2000EP-0003145.

XX PR 02-FEB-2001; 2001EP-0001449.

XX XX

XX PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;

XX DR WPI: 2001-550253/61.

XX PT A library of DNA fragments of Escherichia coli strains for the

XX PT phylogenetic determination of a given strain comprises polynucleotides of

XX PT nature B2/D+ A- -

XX PS Example 6; Fig 6; 646pp; English.

XX CC The invention relates to a library of DNA fragments of Escherichia coli

XX CC strains comprising polynucleotides (ABA89577-ABA89729 and ABA89533)

XX CC and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature

XX CC B2/D+A-. The polynucleotides have potential antiinflammatory,

XX CC antibacterial and immunosuppressive activity as part of pharmaceutical

XX CC compositions used to treat, palliate or prevent extra-intestinal E. coli

XX CC infections. The polypeptides are useful for determining the phylogenetic

XX CC group of a given E. coli strain. These polypeptides can detect and treat

XX CC an undesired development of E. coli, particularly an extra-intestinal

XX CC infection that include systemic and non-diarrhoeal infections such as

XX CC septicemia, pyelonephritis and meningitis this is particularly

XX CC advantageous as bacterial resistance is increasing with the more

XX CC frequent use of broad spectrum antibiotics.

XX SQ Sequence 356 AA;

Query Match 11.5%; Score 146; DB 22; Length 356;

Best Local Similarity 24.4%; Pred. No. 2.6e-07;

Matches 87; Conservative 40; Mismatches 92; Indels 138; Gaps 18;

QY 11 LVISGGAAGIAHIGVLKALNEIGIRVRLSGVSGAIVSVYVYAGVSP-----60
 DB 9 LVISGGGAVAYAGVAVKALAECCGQISWVSGASIGAFNGAIIAA--SPDSEAAVRLTA 66
 QY 61 -----GMSILKRVNMLKFKK-----PPLKG-----LIGW 87
 DB 67 LMDHLGKMOVLAVRVLVYFESLK-----KLFQAMNLCOIFGRAGALTLTLRIHISLNGF 121
 QY 88 EKAIKRLBEVLVPRRIEKLK--IPYICATDLYSGRALYLS-----ECSL-----130
 DB 122 DNLM-----AQPLISDEPLTALDHTLDADLADGIPLYVSLYPTGGMQDIIIDICIRAE 176
 QY 131 -----IP-----ALGSCAIPGIFEPVYKYNLLVNDG-----IYNLL 163
 DB 177 GAGTTNAVFOHTQSLEPRGQKALALASALPLLRPREVGCTMGDGMGRANMGNT 236
 QY 164 PVEPQSGIPYCV-----DVLPISEKDIK-----NIIHLIRSP 200
 DB 237 PVTEPLVDAGCMVTVHLSLMDRQAFPDYTLIEIRPRKRLKXAGDGNMGGLT--SF 294
 QY 201 FLAVRSSEKKEKFCDLVIVPELEPFPDVRK--ADOIMERGYI--KALEVLS 251
 DB 295 TLA-HTDAMRQCGEDMLTME-----HIKPIAAQVLSRSETVLQSLERTE 343

RESULT 13

AY38777
 ID AAY38777 standard; Protein; 149 AA.

XX AAY38777;
 AC

XX 08-OCT-1999 (first entry)
 DT

DE Neisseria meningitidis antigen encoded by a partial ORF137.
 XX

KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 treatment; Neisseria infection; meningitis; septicemia; gonorrhea.
 XX

OS Neisseria meningitidis.
 XX

PN W09924578-A2.
 XX

PD 20-MAY-1999.
 XX

PF 09-OCT-1998; 98WO-IB01665.
 XX

PR 01-SEP-1998; 98GB-0019016.
 XX

PR 06-NOV-1997; 97GB-0023516.
 XX

PR 14-NOV-1997; 97GB-0024190.
 XX

PR 18-NOV-1997; 97GB-0024386.
 XX

PR 27-NOV-1997; 97GB-0025158.
 XX

PR 10-DEC-1997; 97GB-0026147.
 XX

PR 14-JAN-1998; 98GB-0000759.
 XX

PA (CHIR-) CHIRON SPA.
 XX

PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
 XX

DR WPI; 1999-327407/27.
 XX

DR N-PSDB; AAZ1212.
 XX

XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
 PT diagnosis, treatment and prevention of infection
 PT

PS Claim 4; Page 322; 524pp; English.
 CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
 CC and N. gonorrhoeae antigenic proteins. They are encoded by open
 CC reading frames (ORFs) AAZ1972-21258. The antigenic proteins,
 CC their fragments, their nucleic acids and antibodies are used for
 CC diagnosis, prevention (as vaccines) or treatment of Neisseria
 CC infections, such as meningitis, septicemia and gonorrhea. Both

CC organisms are closely related. Fragments of the nucleic acids
 CC are useful as hybridisation probes and antisense reagents.

XX Sequence 149 AA;
 SQ

Query Match 11.4%; Score 145; DB 20; Length 149;
 Best Local Similarity 50.0%; Pred. No. 9.5e-08;
 Matches 26; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 9 INLVSGGAAGIAHIGVLKALNEIGIRVRLSGVSGAIVSVYVYAGVSP 60
 DB 46 VGLALGGASKGPAHVGIKVLKENGIVKVTGTSAGSIVGNLFASGMSPD 97

RESULT 14

ABP30534
 ID ABP30534 standard; Protein; 269 AA.

XX ABP30534;
 AC

XX 02-JUL-2002 (first entry)
 DT

DE Streptococcus polypeptide SEQ ID NO 10244.
 XX

KM Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;
 group A streptococcus; Streptococcus pyogenes; antibacterial;
 XX

KM antinflammatory; infection; vaccine; meningitis; gene therapy.
 XX

OS Streptococcus agalactiae.
 XX

PN W0200234771-A2.
 XX

PD 02-MAY-2002.
 XX

PF 29-OCT-2001; 2001WO-GB04789.
 XX

PR 27-OCT-2000; 2000GB-0026333.
 XX

PR 24-NOV-2000; 2000GB-0028727.
 XX

PR 07-MAR-2001; 2001GB-0005640.
 XX

PA (CHIR-) CHIRON SPA.
 XX

PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Frazer C;
 XX

PI Tectelin H;
 XX

DR WPI; 2002-352536/38.
 XX

DR N-PSDB; ABA71165.
 XX

PS Claim 1; Page 4155; 4525pp; English.
 XX

XX The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and anti-inflammatory
 CC activity. (1), nucleic acids encoding (1), ABA66044-ABA71526 and
 CC antibodies that bind (1) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a
 CC biological sample. (1) is used to determine whether a compound binds to
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (1) may be used to recombinantly produce (1) and may be
 CC used in gene therapy. Antibodies to (1) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.

Thu Jun 19 17:30:42 2003

us-09-903-410-36.rat

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 19, 2003, 15:01:30 ; Search time 23 Seconds

(without alignments)
321.094 Million cell updates/sec

Title: US-09-903-410-36

Perfect score: 1273

Sequence: 1 LRKRKEEINLVLSGAAKG.....RKADQWNGVYIKALVLS 251

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*

2: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*

3: /cgn2_6/prodata/1/1aa/6A.COMB.pep:*

4: /cgn2_6/prodata/1/1aa/6B.COMB.pep:*

5: /cgn2_6/prodata/1/1aa/ECTUS.COMB.pep:*

6: /cgn2_6/prodata/1/1aa/Dackfilest.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1265	99.4	251	2	US-08-602-359A-36
2	901.5	70.8	249	2	US-08-602-359A-39
3	86.5	6.8	1112	4	US-09-353-585-2
4	84.5	6.6	1112	4	US-09-353-585-3
5	81.5	6.4	393	2	US-08-977-554-2
6	81.5	6.4	393	2	US-09-225-967-2
7	81.5	6.4	393	4	US-09-227-806-2
8	81	6.4	832	1	US-07-977-434-12
9	81	6.4	832	1	US-08-458-819-12
10	81	6.4	832	5	PCT-US91-07035-12
11	78	6.1	488	4	US-09-099-053-2
12	77.5	6.1	408	3	US-09-074-912-2
13	77.5	6.1	408	3	US-09-074-912-4
14	77.5	6.1	408	4	US-09-290-136-2
15	77.5	6.1	408	4	US-09-290-136-4
16	76.5	6.0	984	2	US-08-673-789-6
17	76	6.0	350	2	US-08-852-481-2
18	74.5	5.9	357	4	US-09-193-191-2
19	74.5	5.9	951	1	US-08-162-809-2
20	74	5.8	444	4	US-09-413-814-95
21	74	5.8	444	4	US-09-413-814-96
22	74	5.8	537	4	US-09-134-001C-4091
23	73	5.7	545	4	US-08-687-590-26
24	73	5.7	843	4	US-09-361-631-2
25	73	5.7	3782	4	US-09-105-537-4
26	72.5	5.7	605	1	US-08-485-718-9
27	72.5	5.7	605	2	US-08-484-530-55

28	72.5	5.7	605	2	US-08-827-618A-55	Sequence 55, Appl
29	72.5	5.7	605	3	US-08-483-952A-55	Sequence 55, Appl
30	72.5	5.7	605	4	US-08-476-501-55	Sequence 55, Appl
31	72.5	5.7	662	4	US-09-134-001C-4074	Sequence 4074, Ap
32	72	5.7	673	4	US-08-186-387-8	Sequence 8, Appl
33	72	5.7	718	3	US-09-090-808-2	Sequence 2, Appl
34	72	5.7	718	4	US-09-447-453-2	Sequence 2, Appl
35	72	5.7	740	4	US-09-323-872A-25	Sequence 25, Appl
36	72	5.7	949	4	US-09-196-387-10	Sequence 10, Appl
37	72	5.7	1327	4	US-09-196-387-2	Sequence 2, Appl
38	71.5	5.6	269	4	US-09-355-166-11	Sequence 11, Appl
39	70	5.5	451	4	US-09-412-102-8	Sequence 8, Appl
40	70	5.5	451	4	US-09-217-787-8	Sequence 8, Appl
41	69.5	5.5	397	3	US-08-135-782-2	Sequence 2, Appl
42	69.5	5.5	554	3	US-08-904-871-1	Sequence 2, Appl
43	69.5	5.5	831	2	US-08-677-734A-11	Sequence 11, Appl
44	69.5	5.5	831	4	US-09-097-053-11	Sequence 11, Appl
45	69	5.4	245	4	US-09-369-364A-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1

US-08-602-359A-36

Sequence 36, Application US/08602359A

Patent No. 5942430

GENERAL INFORMATION:

APPLICANT: ROBERTSON, Daniel E.

APPLICANT: MURPHY, Dennis

APPLICANT: REID, John

APPLICANT: MAFFA, Anthony

APPLICANT: LINK, Steven

APPLICANT: SWANSON, Ronald V.

APPLICANT: WARREN, Patrick V.

APPLICANT: KOSMOTKA, Anna

TITLE OF INVENTION: ESTERASES

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: FISH & RICHARDSON P.C.

STREET: 4225 EXECUTIVE SQUARE, STE 1400

CITY: LA JOLLA

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH DISKETTE

COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS

SOFTWARE: WORD PERFECT 6.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/602,359A

FILING DATE: February 16, 1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: HALL, LISA A.

REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: 09010/010001

TELEPHONE: 619-678-5099

TELEFAX: 619-678-5099

INFORMATION FOR SEQ ID NO: 36:

SEQUENCE CHARACTERISTICS:

LENGTH: 251 AMINO ACIDS

TYPE: AMINO ACIDS

TOPOLOGY: LINEAR

MOLECULE TYPE: PROTEIN

US-08-602-359A-36

Query Match 99.4%, Score 1265, DB 2, Length 251;

MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Tomato
STRAIN: Cf2
SEQUENCE DESCRIPTION: SEQ ID NO: 2
US-09-353-585-2

Query Match 6.8%; Score 86.5; DB 4; Length 1112;
Best Local Similarity 30.3%; Pred. No. 1.2;
Matches 46; Conservative 25; Mismatches 56; Indels 25; Gaps 10;

QY 3 LRKEERINLV---ISGGAAGIAHIGVKAINEIGIRVALSG---VSAGAI--VSVPY- 53
DB LNNLSRLNLTNNQSGSIPE---EIGYLRSLNDLGLSENALNGSIIPASIGNNLSMTLY 606
QY 54 ----AGSYSPGMSILKRVNMLKFKRPPLKGLIGWEKA-IRPLEVLPYRRIEKL 108
DB 607 YNNQSGSIPEEI-GYLSLTYLSL--GNNSLNGLLPASFPAMRMQLIINDNNLIGE 663
QY 109 PTTCATDLYSGRALYISGSL---IPALGS 137
DB 664 PSSVC--NLTSLVLYMPRNMLKGVPCQCGN 693

RESULT 4

US-09-353-585-3
Sequence 3, Application US/09353585
Patent No. 6287865

GENERAL INFORMATION:

APPLICANT: Dixon, Mark S
Jones, David A
Jones, Jonathan DG
TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye PC
STREET: 8th Floor, 1100 No. 6287865th Glebe Road

CITY: Arlington
STATE: Virginia

COUNTRY: United States of America
ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/353,585
FILING DATE: 15-Jul-1999

CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12Q 1/68
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/930,277
FILING DATE: 27-OCT-1997

APPLICATION NUMBER: PCT/GB96/00785
FILING DATE: 01-APR-1996
APPLICATION NUMBER: GB 950658.5
FILING DATE: 31-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Ms Mary J Wilson
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 620-69

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 1112 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear

MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Tomato
STRAIN: Cf2
SEQUENCE DESCRIPTION: SEQ ID NO: 3
US-09-353-585-3

Query Match 6.6%; Score 84.5; DB 4; Length 1112;
Best Local Similarity 27.9%; Pred. No. 2;
Matches 46; Conservative 23; Mismatches 45; Indels 51; Gaps 12;

QY 3 LRKEERINLV---ISGGAAGIAHIGVKAINEIGIRVALSG---VSAGAI--VSVPY- 53
DB LNNLSRLNLTNNQSGSIPE---EIGYLRSLNDLGLSENALNGSIIPASIGNNLSMTLY 606
QY 54 ----AGSYSPGMSILKRVNMLKFKRPPLKGLIGWEKAIRPLEVLP---YRIE 104
DB 607 YNNQSGSIPEEI-GYLSLTYLSL--GNNSLNGLLPASFPAMRMQLIINDNNLIGE 650
QY 105 KL-----EPTTCATDLYSGRALYISGSL---IPALGS 137
DB 651 ALIINDNNLIGRIPSSVC--NLTSLVLYMPRNMLKGVPCQCGN 693

RESULT 5

US-08-977-554-2
Sequence 2, Application US/08977554
Patent No. 5891672

GENERAL INFORMATION:

APPLICANT: Palmer, Leslie M.
APPLICANT: Pedon, Jason C.
APPLICANT: Warren, Richard L.
APPLICANT: Train, Christopher M.

APPLICANT: Wang, Min

APPLICANT: Uworeki, Deborah D.
APPLICANT: Mooney, Jeffrey

APPLICANT: Debonck, Christine
APPLICANT: Zhong, Yiyi

APPLICANT: Black, Michael
TITLE OF INVENTION: r1ba

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert, Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre

CITY: Philadelphia
STATE: PA
COUNTRY: USA

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/977,554
FILING DATE:

CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/02318
FILING DATE: 19-FEB-1997

ATTORNEY/AGENT INFORMATION:
NAME: Dickinson, O. Todd
REGISTRATION NUMBER: 28,354
REFERENCE/DOCKET NUMBER: P50444-07

TELECOMMUNICATION INFORMATION:
TELEPHONE: 215/994-2252
TELEFAX: 215/994-2222

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 393 amino acids
TYPE: amino acid

```
;
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-977-554-2

Query Match
Best Local Similarity 6.4%; Score 81.5; DB 2; Length 393;
Matches 58; Conservative 39; Mismatches 91; Indels 75; Gaps 13;

QY 32 ELGIRVALSGVS-AGAIVSVFYASGSPGMSLLKRVNWLKLFKFKPPLKGLIGNWEKA 90
DB 144 EAAVDLAKLTGAKPAGVCEIMNDGDTMAKGOD-----LQNFKEKHOLKMIT----- 190
QY 91 IRLEEVLPYRR-----IE---KLEIPYICATDLYSGRALYLSGSLI----- 131
DB 191 ---IDDIIEYRKLEPEIEFKAKVMPDFTGDMYGFKATYDEEIVVLTGKAIRQHEN 247
QY 132 PALGSCAIPGIF-----EPVEYKN-----YLLVDG---GIVNNLPVEPFQ 169
DB 248 VRLHSACTGDIHFSORCDGQAQLESSMKYINEHGMIIYLPQEGRGIGLNLKRAYELI 307
QY 170 ESGIPTVCVDVLPISPEKDIK-----EPVEYKN-----YLLVDG---GIVNNLPVEPFQ 169
DB 308 EOGDTVTAN-LAIGFDELDYHIAAQILKYFNIHINLLS---NNPSKPEGLKQY-GI 362
QY 218 VIVPELEBFTPLDVRKADQIMER 240
DB 363 DIAERIEVIVPETVHNHDYMTK 385

RESULT 6
US-09-225-967-2
; Sequence 2, Application US/09225967
; Patent No. 6171598
; GENERAL INFORMATION:
; APPLICANT: Palmer, Leslie M.
; APPLICANT: Fedon, Jason C.
; APPLICANT: Warren, Richard L.
; APPLICANT: Traini, Christopher M.
; APPLICANT: Wang, Min
; APPLICANT: Jaworski, Deborah D.
; APPLICANT: Mooney, Jeffrey
; APPLICANT: Debouck, Christine
; APPLICANT: Zhong, Yiyi
; APPLICANT: Black, Michael
; TITLE OF INVENTION: riba
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; APPLICATION NUMBER: US/09/225,967
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/977,554
; FILING DATE:
; APPLICATION NUMBER: PCT/US97/02318
; FILING DATE: 19-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Q. Todd
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: P50444-07
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215/994-2252
```

```
;
; TELEFAX: 215/994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 393 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-225-967-2

Query Match
Best Local Similarity 6.4%; Score 81.5; DB 4; Length 393;
Matches 58; Conservative 39; Mismatches 91; Indels 75; Gaps 13;

QY 32 ELGIRVALSGVS-AGAIVSVFYASGSPGMSLLKRVNWLKLFKFKPPLKGLIGNWEKA 90
DB 144 EAAVDLAKLTGAKPAGVCEIMNDGDTMAKGOD-----LQNFKEKHOLKMIT----- 190
QY 91 IRLEEVLPYRR-----IE---KLEIPYICATDLYSGRALYLSGSLI----- 131
DB 191 ---IDDIIEYRKLEPEIEFKAKVMPDFTGDMYGFKATYDEEIVVLTGKAIRQHEN 247
QY 132 PALGSCAIPGIF-----EPVEYKN-----YLLVDG---GIVNNLPVEPFQ 169
DB 248 VRLHSACTGDIHFSORCDGQAQLESSMKYINEHGMIIYLPQEGRGIGLNLKRAYELI 307
QY 170 ESGIPTVCVDVLPISPEKDIK-----EPVEYKN-----YLLVDG---GIVNNLPVEPFQ 169
DB 308 EOGDTVTAN-LAIGFDELDYHIAAQILKYFNIHINLLS---NNPSKPEGLKQY-GI 362
QY 218 VIVPELEBFTPLDVRKADQIMER 240
DB 363 DIAERIEVIVPETVHNHDYMTK 385

RESULT 7
US-09-227-806-2
; Sequence 2, Application US/09227806
; Patent No. 6280971
; GENERAL INFORMATION:
; APPLICANT: Palmer, Leslie M.
; APPLICANT: Fedon, Jason C.
; APPLICANT: Warren, Richard L.
; APPLICANT: Traini, Christopher M.
; APPLICANT: Wang, Min
; APPLICANT: Jaworski, Deborah D.
; APPLICANT: Mooney, Jeffrey
; APPLICANT: Debouck, Christine
; APPLICANT: Zhong, Yiyi
; APPLICANT: Black, Michael
; TITLE OF INVENTION: riba
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; APPLICATION NUMBER: US/09/227,806
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/977,554
; FILING DATE:
; APPLICATION NUMBER: PCT/US97/02318
; FILING DATE: 19-FEB-1997
```

ATTORNEY/AGENT INFORMATION:
 NAME: Dickinson, O. Todd
 REGISTRATION NUMBER: 28,354
 REFERENCE/DOCKET NUMBER: P50444-07
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215/994-2252
 TELEFAX: 215/994-2222
 TELEX:
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 393 amino acids
 TYPE: amino acid
 STRANDBENESS: single
 TOPOLOGY: linear
 US-09-227-806-2

Query Match 6.4%; Score 81.5; DB 4; Length 393;
 Best Local Similarity 22.1%; Pred. No. 0.88; Mismatches 91; Indels 75; Gaps 13;
 Matches 58; Conservative 39;

32 ELGIRVRLSGVS-AGAIIVSVFASGYSPEGMFSLKXVNMKLFKFKPLKGLIGKEKA 90
 144 EAAVDLAKLTGAKPAGVIGCEIMNDGTMAGQD-----LQNKKEGQLKMIT----- 190
 QY 91 IRPLEEVLPRR-----IE---KLEPTTCATDLYSGRALYLSGSLI----- 131
 DB 191 ---IDDLIEYRKLEPELEFKAKVKMPTGTFDMYGFKATYTBELIVLTGKALROHEN 247
 QY 132 PALGSCAIPGIF-----BPVEYKN-----YLLVQG-----GIYNNLPVEPFQ 169
 DB 248 VRLHSACLTGDIIFISQRCDCGAQLSSMKYINHGAMITYLPQERGGIGLTKRAVELI 307
 QY 170 ESGIPTCVVLPFPEPEKDK-----NIHILTSFFLAIVNSSEKKEPCDL 217
 DB 308 EGGIDVTYAN-LALGPFEDLRADYHIAQLTKYFNIEHINLS--NNSKKEGAKQY-GI 362
 QY 218 VIVPELEBFTPLDYRKADQIMER 240
 DB 363 DIARIEVIVPEYVHNHDVWVTK 385

RESULT 8

US-07-977-434-12
 Sequence 12, Application US/07977434
 Patent No. 5466591.
 GENERAL INFORMATION:
 APPLICANT: Gelfand, David H.
 APPLICANT: Abramson, Richard D.
 TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF
 NUMBER OF SEQUENCES: 38
 CORRESPONDENCE ADDRESS:
 ADDRESSER: Hoffmann-La Roche Inc.
 STREET: 340 Kingsland Street
 CITY: Nutley
 STATE: New Jersey
 ZIP: 07110-1199
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: Macintosh
 OPERATING SYSTEM: 7
 SOFTWARE: WordPerfect 2.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/977,434
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 590,490
 FILING DATE: 28-SEP-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 590,466
 FILING DATE: 28-SEP-1990
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 590,213
 FILING DATE: 28-SEP-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 523,394
 FILING DATE: 15-MAY-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 143,441
 FILING DATE: 12-JAN-1988
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 063,509
 FILING DATE: 17-JUN-1987
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 699,241
 FILING DATE: 22-AUG-1986
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 746,121
 FILING DATE: 15-AUG-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/US90/07641
 FILING DATE: 21-DEC-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 585,471
 FILING DATE: 20-SEP-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 455,611
 FILING DATE: 22-DEC-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 609,157
 FILING DATE: 02-NOV-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 557,517
 FILING DATE: 24-JUL-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Iannu Cseart
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: Case No. 5466591 8753
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (510) 814-2972
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 892 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-977-434-12

Query Match 6.4%; Score 81; DB 1; Length 892;
 Best Local Similarity 22.6%; Pred. No. 3.5;
 Matches 53; Conservative 47; Mismatches 75; Indels 60; Gaps 14;

QY 26 VLAINEIGIRVRLSGVSAGAIIVSVFASGYSPEGMFSLKXVNMKLFKFKPLKGLI 65
 DB 95 VEEIVDMLGIRKIKYKLGFEADDIITL-----SKKEESDEKKNV-----IT 136
 QY 86 GMEKAIPLFEELVPRRIKLSIPTVICATDLYSGRALYLSGSLIPALGSCAIPGIFE 145
 DB 137 GDXDLQLVSDKVFWRVR-----GITDL-----VLV-DRNKVIEKY----- 176
 QY 146 PVEKIVLVDDGIVNMLEVPEPESGIPV-----CVDPVPIPEKDKIITL-----HILR 198
 DB 177 PEQFKYILSVGQIDNIP-----GVKGIQKTNVSL--KKNLSLEVNLKINMLIR 227
 QY 199 SFPLAVNSSEKKEPCDLVIVPELEBFTPLDYRKADQIMERG--IKALEVISE 251
 DB 228 KLRRLLEDSKEDLOKSEIELV--ELIYDPMVDER-DEIIRGYNDPKLTKYAKK 278

RESULT 9

US-08-458-819-12
 Sequence 12, Application US/08458819
 Patent No. 5795762
 GENERAL INFORMATION:
 APPLICANT: Gelfand, David H.

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 585,471
FILING DATE: 20-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 455,611
FILING DATE: 22-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 609,157
FILING DATE: 02-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 557,517
FILING DATE: 24-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: SIAA PH.D. STACEY R.
REGISTRATION NUMBER: 37,630
REFERENCE/DOCKET NUMBER: Case No. 2580
TELEPHONE: 415-420-3300
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 892 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US91-07035-12

Query Match 6.4%; Score 81; DB 5; Length 892;
Best Local Similarity 22.6%; Pred. No. 3.5;
Matches 53; Conservative 47; Mismatches 75; Indels 60; Gaps 14;

QY 26 VLKAINELGIRVRLSGVSAGATVSVFVSGVSPGMSLKRNVNMLKLFKPKLKGIL 85
DB 95 VEBIVDNLGKIKVKBEPFADQIATL-----SKKPSDEPKNI-----II 136
QY 86 GMEKAIKRLBEVLPRRIEGLIPTICATDIYSGRALYSEGLIPALGSCAIPGIFE 145
DB 137 GDKKLLVSDSKVPMVWER-----GITDL-----VLY-DRNKVIKY-----GIY- 176
QY 146 PVEKYNILVDGIVNNLPVEPQESGIPV-----CVDYLPTEPEKQIKNII--HILR 198
DB 177 PEQPKDYLISLVDQIDNIF-----GVKGIKKTAVSL--KKNSLENNLKNINILTE 227
QY 199 SFLAVNSSEKRFCDLYIVPELEFPLVYRKADQIMERY--IKALEVLSB 251
DB 228 KLRLILDSKEDLQKSIETV--ELIYDVPMDEK-DELIYGYNPDLLKVLKK 278

RESULT 11

US-09-099-053-2
Sequence 2, Application US/09099053
Patent No. 6388063
GENERAL INFORMATION:
APPLICANT: Greg Plowman
APPLICANT: Susan Onrust
APPLICANT: David Markby
APPLICANT: Sara Courtneidge
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF
NUMBER OF INVENTIONS: SAD RELATED DISORDERS
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Filth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FastSeq for Windows 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/099,053
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/049,914
FILING DATE: June 18, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 235/121
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELFX: 67-3510
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 488 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-099-053-2

Query Match 6.1%; Score 78; DB 4; Length 488;
Best Local Similarity 21.9%; Pred. No. 3.1;
Matches 42; Conservative 27; Mismatches 75; Indels 48; Gaps 6;

QY 6 FEEINVLISGAKAKIHAIGV-----LKAINELG--IRVRLSGVAGATVGVFVAGYS 58
DB 52 FPGFLATVDFPARGGELSFRAGDRLCALREGGYIFARRISGQSAGLVPTHTVAAS 111
QY 59 PEM-----ESLIRVNMMLK-----PKRPLKGLGMEKAIKRLBEVLPRR 102
DB 112 PETLSDQPFYFGVRRTOAQLLSPNREAFIRSESSLGYSLSVRAQAKVCHR- 170
QY 103 IETKILPTICATDIYSGRALYSEGLIPALGSCAIPGIFEPVEKYNILVDGIVNN 162
DB 171 -----VSMADGSLYLQGRLEPGL-----EBLLTYKANNWILION 206
QY 163 LPVEPQESGP 174
DB 207 PLIQPCMPQKAP 218

RESULT 12

US-09-074-912-2
Sequence 2, Application US/09074912
Patent No. 6057491
GENERAL INFORMATION:
APPLICANT: CIGAN, AMY L.
APPLICANT: CZAPLA, THOMAS H.
APPLICANT: FALLIS, LYNN
APPLICANT: MEYER, TERRY E.
APPLICANT: MINDELL, SCOTT A.
APPLICANT: SABUS, BRIAN
APPLICANT: SCHUBERT, KAREL
TITLE OF INVENTION: PROTEINS HAVING INSECTICIDAL ACTIVITIES
NUMBER OF INVENTIONS: AND METHODS OF USE
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: W. MORRAY SPRUILL (ALSTON & BIRD, LLP)
STREET: 3605 GLENWOOD AVE.
CITY: RALEIGH
STATE: NC
COUNTRY: USA
ZIP: 27622
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

Thu Jun 19 17:30:42 2003

us-09-903-410-36.ra1

Page 9

```
DB      151 NETLDDQTTNVVIPSF---DILKLRPTIFSTFKLEVEPELNVKLSDVCMGTSAPYIF 206
QY      145 EPEVEYKN---YLLVDGGIVNNLPEVPPQESGIPVVCVDVLPLEPEKDIKNIL 193
DB      207 PPEYFKHSGTBERNLVDGAILADIIPA-----PVALSEVILQOEKYN-KEIL 250
```

RESULT 15

```
US-09-290-136-4
; Sequence 4, Application US/09290136
; Patent No. 6339144
; GENERAL INFORMATION:
; APPLICANT: Cigan, Amy L.
; APPLICANT: Czapl, Thomas H.
; APPLICANT: Fajlis, Lynn
; APPLICANT: Meyer, Terry E.
; APPLICANT: Mundell, Scott A.
; APPLICANT: Sabus, Brian
; APPLICANT: Schubert, Karel
; TITLE OF INVENTION: Proteins Having Insecticidal Activities and Methods of
; TITLE OF INVENTION: Use
; FILE REFERENCE: 5718-9A, 035718/180486
; CURRENT APPLICATION NUMBER: US/09/290.136
; CURRENT FILING DATE: 1999-04-13
; EARLIER APPLICATION NUMBER: 09/074,912
; EARLIER FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 408
; TYPE: PRT
; ORGANISM: Pentaclethra macroloba
US-09-290-136-4
```

```
Query Match      6.1%; Score 77.5; DB 4; Length 408;
Best Local Similarity 21.9%; Pred. No. 2.6;
Matches 51; Conservative 34; Mismatches 77; Indels 71; Gaps 12;
```

```
QY      11 LVLSGGAAGKGINHIGVLKAIN-----ELGIRVRALSGVAGAIIVSVFYASG----- 56
DB      39 LAIDGGGIRGILPGVILKQLENTLQRMWSSARLAEYFDVAVAGSTGGIITAILTADPQN 98
QY      57 -----YSPEGMF-----SLKRVNMLKU-FKRPPLKGLIGMEKAIREFLEVDLYRR 102
DB      99 KDRPLVAAEEIIDFYIEHGPSIFPNKSTACSLPGIFCPKRYDG-----KYLOEIIISOKL 150
QY      103 IEKL-----EIPFYICATDLYSGRALYLSGSL--IPAL-----LGSCAIPGIF 144
DB      151 NETLDDQTTNVVIPSF---DILKLRPTIFSTFKLEVEPELNVKLSDVCMGTSAPYIF 206
QY      145 EPEVEYKN---YLLVDGGIVNNLPEVPPQESGIPVVCVDVLPLEPEKDIKNIL 193
DB      207 PPEYFKHSGTBERNLVDGAILADIIPA-----PVALSEVILQOEKYN-KEIL 250
```

Search completed: June 19, 2003, 15:05:27
Job time : 24 secs

Thu Jun 19 17:30:42 2003

US-09-903-410-36.rapb

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

CM protein - protein search, using sw model

Run on: June 19, 2003, 15:04:11 ; Search time 182 Seconds
(without alignments)
149.230 Million cell updates/sec

Title: US-09-903-410-36

Sequence: 1 LRLKRFERINVLISGGAAGK.....RKADQIMRGYKALEVLSR 251

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 10820613 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEM_PUB pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEM_PUB pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEM_PUB pep:*
4: /cgn2_6/ptodata/1/pubpaa/US07_NEM_PUB pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep:*
8: /cgn2_6/ptodata/1/pubpaa/US09_NEM_PUB pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEM_PUB pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEM_PUB pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1273	100.0	251	10	US-09-903-410-36 Sequence 36, Appl
2	1265	99.4	251	9	US-10-027-805-36 Sequence 36, Appl
3	1265	99.4	251	9	US-10-027-804-36 Sequence 36, Appl
4	913.5	71.8	249	10	US-09-903-410-39 Sequence 39, Appl
5	901.5	70.8	249	9	US-10-027-805-39 Sequence 39, Appl
6	901.5	70.8	249	9	US-10-027-804-39 Sequence 39, Appl
7	139.5	11.0	425	9	US-10-147-026-10 Sequence 10, Appl
8	109.5	8.6	504	10	US-09-765-205-40 Sequence 40, Appl
9	86.5	6.7	3354	9	US-10-034-937-14 Sequence 14, Appl
10	86.5	6.7	3354	9	US-10-160-758-11 Sequence 11, Appl
11	85.5	6.6	3354	9	US-10-034-937-12 Sequence 12, Appl
12	83.5	6.6	330	9	US-10-034-937-10 Sequence 10, Appl
13	83.5	6.6	455	9	US-10-082-627A-25 Sequence 25, Appl
14	83	6.5	391	10	US-09-815-242-5379 Sequence 5379, A
15	83	6.5	393	10	US-09-815-242-12285 Sequence 12285, A
16	81.5	6.4	390	9	US-10-034-937-12 Sequence 12, Appl
17	81.5	6.4	390	9	US-10-034-937-12 Sequence 12, Appl
18	81	6.4	802	9	US-10-245-103-78 Sequence 78, Appl
19	81	6.4	802	9	US-10-245-107-78 Sequence 78, Appl

20	81	6.4	802	9	US-10-245-143-78	Sequence 78, Appl
21	81	6.4	802	9	US-10-245-771-78	Sequence 78, Appl
22	81	6.4	802	9	US-10-245-851-78	Sequence 78, Appl
23	81	6.4	802	9	US-10-245-883-78	Sequence 78, Appl
24	81	6.4	802	9	US-10-237-535-78	Sequence 78, Appl
25	81	6.4	802	9	US-10-338-183-78	Sequence 78, Appl
26	81	6.4	802	9	US-10-338-283-78	Sequence 78, Appl
27	81	6.4	802	9	US-10-338-370-78	Sequence 78, Appl
28	81	6.4	802	9	US-10-245-055-78	Sequence 78, Appl
29	81	6.4	802	9	US-10-245-147-78	Sequence 78, Appl
30	81	6.4	802	9	US-10-245-730-78	Sequence 78, Appl
31	81	6.4	802	9	US-10-245-739-78	Sequence 78, Appl
32	81	6.4	802	9	US-10-246-210-78	Sequence 78, Appl
33	81	6.4	802	9	US-10-239-196-78	Sequence 78, Appl
34	81	6.4	802	9	US-10-243-024-78	Sequence 78, Appl
35	81	6.4	802	9	US-10-243-409-78	Sequence 78, Appl
36	81	6.4	802	9	US-10-245-033-78	Sequence 78, Appl
37	81	6.4	802	9	US-10-245-621-78	Sequence 78, Appl
38	81	6.4	802	9	US-10-245-880-78	Sequence 78, Appl
39	81	6.4	802	9	US-10-243-095-78	Sequence 78, Appl
40	81	6.4	802	9	US-10-245-185-78	Sequence 78, Appl
41	81	6.4	802	9	US-10-245-427-78	Sequence 78, Appl
42	81	6.4	802	9	US-10-245-473-78	Sequence 78, Appl
43	81	6.4	802	9	US-10-245-770-78	Sequence 78, Appl
44	81	6.4	802	9	US-10-245-877-78	Sequence 78, Appl
45	81	6.4	802	9	US-10-246-976-78	Sequence 78, Appl

ALIGNMENTS

RESULT 1

US-09-903-410-36
Sequence 36, Application US/09903410
Patent No. US20020146799A1
GENERAL INFORMATION:
APPLICANT: DIVERSA CORPORATION
APPLICANT: ROBERTSON, Dan
APPLICANT: MURPHY, Dennis
APPLICANT: REID, John
APPLICANT: MAREIA, Anthony
APPLICANT: LINK, Steven
APPLICANT: SWANSON, Ronald
APPLICANT: WARREN, Patrick
APPLICANT: KOSMOTKA, Anna
TITLE OF INVENTION: ENZYMS HAVING ESTERASE ACTIVITY AND METHODS OF USE THEREOF
FILE REFERENCE: DIVERS1180-2
CURRENT APPLICATION NUMBER: US/09/903,410
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: US 09/382,242
PRIOR FILING DATE: 1999-08-24
PRIOR APPLICATION NUMBER: US 08/602,359
PRIOR FILING DATE: 1996-02-16
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn version 3.0
SEQ ID NO 36
LENGTH: 251
TYPE: PRT
ORGANISM: Aquifex pyrophilus
US-09-903-410-36

Query Match 100.0%; Score 1273; DB 10; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.2e-120;
Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	LRLKRFERINVLISGGAAGKIAHIGVKAINEGIRVRLSGVSAGAIYSVYAGSYSE	60
DB	1	LRLKRFERINVLISGGAAGKIAHIGVKAINEGIRVRLSGVSAGAIYSVYAGSYSE	60
QY	61	GMFSLKRVVWLKLFKFKPPKLGKIGWEKARIFLEBEVLPYRRIEKLIEPTTICATDIVSG	120
DB	61	GMFSLKRVVWLKLFKFKPPKLGKIGWEKARIFLEBEVLPYRRIEKLIEPTTICATDIVSG	120

Qy 121 RALYSEGLIPALLGSCAIGIEPEVEYKYNLLVDGGIYVNNLPVPPFOESGIPVCDV 180
 Db 121 RALYSEGLIPALLGSCAIGIEPEVEYKYNLLVDGGIYVNNLPVPPFOESGIPVCDV 180
 Qy 181 LPIEPEKDKNLLHLLRSFFLAVERNSSEKKEFCDLIVIVPELEETPLDVRKADQIMER 240
 Db 181 LPIEPEKDKNLLHLLRSFFLAVERNSSEKKEFCDLIVIVPELEETPLDVRKADQIMER 240
 Qy 241 GYIKALEVLS 251
 Db 241 GYIKALEVLS 251

RESULT 2

US-10-027-805-36
 ; Sequence 36, Application US/10027805
 ; Patent No. US20020164725A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ROBERTSON, Daniel E.
 ; MURPHY, Dennis
 ; REID, John
 ; MAFFIA, Anthony
 ; LINK, Steven
 ; SWANSON, Ronald V.
 ; WARREN, Patrick V.
 ; KOSMOTKA, Anna
 ; TITLE OF INVENTION: ESTERASES
 ; NUMBER OF SEQUENCES: 42
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: FISH & RICHARDSON P.C.
 ; STREET: 4225 EXECUTIVE SQUARE, STE 1400
 ; CITY: LA JOLLA
 ; STATE: CALIFORNIA
 ; COUNTRY: USA
 ; ZIP: 92037
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 INCH DISKETTE
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: WORD PERFECT 6.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/027.805
 ; FILING DATE: 21-Dec-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/602,359
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: HAILE, LISA A.
 ; REFERENCE/DOCKET NUMBER: 09010/010001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 619-678-5070
 ; TELEFAX: 619-678-5099
 ; INFORMATION FOR SEQ ID NO: 36:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 251 AMINO ACIDS
 ; TYPE: AMINO ACID
 ; TOPOLOGY: LINEAR
 ; MOLECULE TYPE: PROTEIN
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 36:
 US-10-027-805-36

Query Match 99.4%; Score 1265; DB 9; Length 251;
 Best Local Similarity 99.6%; Pred. No. 7.6e-120;
 Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 LRLRKEEINLVLSGGAAGIAHIGVLKAINELGIRVRLSGVSGAIVSVFYASGYSPE 60
 Db 1 LRLRKEEINLVLSGGAAGIAHIGVLKAINELGIRVRLSGVSGAIVSVFYASGYSPE 60
 Qy 61 GMPSLLKRVNMLKFKPKPLKGLIGWEKAIREFLEEVLPYRRIEKLIPTTYICATDLYSG 120

Db 61 GMPSLLKRVNMLKFKPKPLKGLIGWEKAIREFLEEVLPYRRIEKLIPTTYICATDLYSG 120
 Qy 121 RALYSEGLIPALLGSCAIGIEPEVEYKYNLLVDGGIYVNNLPVPPFOESGIPVCDV 180
 Db 121 RALYSEGLIPALLGSCAIGIEPEVEYKYNLLVDGGIYVNNLPVPPFOESGIPVCDV 180
 Qy 181 LPIEPEKDKNLLHLLRSFFLAVERNSSEKKEFCDLIVIVPELEETPLDVRKADQIMER 240
 Db 181 LPIEPEKDKNLLHLLRSFFLAVERNSSEKKEFCDLIVIVPELEETPLDVRKADQIMER 240
 Qy 241 GYIKALEVLS 251
 Db 241 GYIKALEVLS 251

RESULT 3

US-10-027-804-36
 ; Sequence 36, Application US/10027804
 ; Publication No. US20030054530A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ROBERTSON, Daniel E.
 ; MURPHY, Dennis
 ; REID, John
 ; MAFFIA, Anthony
 ; LINK, Steven
 ; SWANSON, Ronald V.
 ; WARREN, Patrick V.
 ; KOSMOTKA, Anna
 ; TITLE OF INVENTION: ESTERASES
 ; NUMBER OF SEQUENCES: 42
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: FISH & RICHARDSON P.C.
 ; STREET: 4225 EXECUTIVE SQUARE, STE 1400
 ; CITY: LA JOLLA
 ; STATE: CALIFORNIA
 ; COUNTRY: USA
 ; ZIP: 92037
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 INCH DISKETTE
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: WORD PERFECT 6.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/027.804
 ; FILING DATE: 21-Dec-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/602,359
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: HAILE, LISA A.
 ; REFERENCE/DOCKET NUMBER: 09010/010001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 619-678-5070
 ; TELEFAX: 619-678-5099
 ; INFORMATION FOR SEQ ID NO: 36:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 251 AMINO ACIDS
 ; TYPE: AMINO ACID
 ; TOPOLOGY: LINEAR
 ; MOLECULE TYPE: PROTEIN
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 36:
 US-10-027-804-36

Query Match 99.4%; Score 1265; DB 9; Length 251;
 Best Local Similarity 99.6%; Pred. No. 7.6e-120;
 Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 LRLRKEEINLVLSGGAAGIAHIGVLKAINELGIRVRLSGVSGAIVSVFYASGYSPE 60
 Db 1 LRLRKEEINLVLSGGAAGIAHIGVLKAINELGIRVRLSGVSGAIVSVFYASGYSPE 60

QY 61 GMSFLKRVNMLKLFKRPPLKGLIGWEKAIPTLEEVLPYRRIEKLPIPTICADLVSG 120
DB 61 GMSFLKRVNMLKLFKRPPLKGLIGWEKAIPTLEEVLPYRRIEKLPIPTICADLVSG 120
QY 121 RALYLSGSLIPALGSCAIPGIFEPVEYKYLVDGIVNNLPVEPFOESGIPYCVDV 180
DB 121 RALYLSGSLIPALGSCAIPGIFEPVEYKYLVDGIVNNLPVEPFOESGIPYCVDV 180
QY 181 LPIPEPKDKNLIHLILSFFLAVNSSEKKEPCDVIYVPELBEFTPLDVRKADQIMER 240
DB 181 LPIPEPKDKNLIHLILSFFLAVNSSEKKEPCDVIYVPELBEFTPLDVRKADQIMER 240
QY 241 GYIKALEVLS 251
DB 241 GYIKALEVLS 251

RESULT 4
US-09-903-410-39
Sequence 39, Application US/09903410
Patent No. US20020146799A1
GENERAL INFORMATION:
APPLICANT: DIVERSA CORPORATION
APPLICANT: ROBERTSON, Dan
APPLICANT: MOREHY, Dennis
APPLICANT: REID, John
APPLICANT: MAFITA, Anthony
APPLICANT: LINK, Steven
APPLICANT: SWANSON, Ronald
APPLICANT: WARREN, Patrick
APPLICANT: KOSMOTKA, Anna
TITLE OF INVENTION: ENZYMS HAVING ESTERASE ACTIVITY AND METHODS OF USE THEREOF
FILE REFERENCE: DIVER1180-2
CURRENT FILING DATE: US/09/903,410
CURRENT APPLICATION NUMBER: 2001-07-10
PRIOR FILING DATE: 1999-08-24
PRIOR APPLICATION NUMBER: US 08/602,359
PRIOR FILING DATE: 1996-02-16
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn version 3.0
SEQ ID NO 39
LENGTH: 249
TYPE: PRT
ORGANISM: Aquifex VF5-34LC
US-09-903-410-39

Query Match 71.8%; Score 913.5; DB 10; Length 249;
Best Local Similarity 72.0%; Pred. No. 2.6e-84;
Matches 177; Conservative 31; Mismatches 37; Indels 1; Gaps 1;
QY 1 LRLKKEEINLVLSGGAAGIAHIGVLAINELGIRVPLSGVSAAGIYVFAASGSP 60
DB 5 LRLKKEEINLVLSGGAAGIAHIGVLAINELGIRVPLSGVSAAGIYVFAASGSP 60
QY 61 GMSFLKRVNMLKLFKRPPLKGLIGWEKAIPTLEEVLPYRRIEKLPIPTICADLVSG 120
DB 61 GMSFLKRVNMLKLFKRPPLKGLIGWEKAIPTLEEVLPYRRIEKLPIPTICADLVSG 120
QY 121 RALYLSGSLIPALGSCAIPGIFEPVEYKYLVDGIVNNLPVEPFOESGIPYCVDV 180
DB 121 RALYLSGSLIPALGSCAIPGIFEPVEYKYLVDGIVNNLPVEPFOESGIPYCVDV 180
QY 181 LPIPEPKDKNLIHLILSFFLAVNSSEKKEPCDVIYVPELBEFTPLDVRKADQIMER 240
DB 181 LPIPEPKDKNLIHLILSFFLAVNSSEKKEPCDVIYVPELBEFTPLDVRKADQIMER 240
QY 241 GYIKAL 246
DB 244 GDMRAL 249

RESULT 5

US-10-027-805-39
Sequence 39, Application US/10027805
Patent No. US20020164725A1
GENERAL INFORMATION:
APPLICANT: ROBERTSON, Daniel E.
APPLICANT: MOREHY, Dennis
APPLICANT: REID, John
APPLICANT: MAFITA, Anthony
APPLICANT: LINK, Steven
APPLICANT: SWANSON, Ronald V.
APPLICANT: WARREN, Patrick V.
APPLICANT: KOSMOTKA, Anna
TITLE OF INVENTION: ESTERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSER: FISH & RICHARDSON P.C.
STREET: 4225 EXECUTIVE SQUARE, STE 1400
CITY: LA JOLLA
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 6.0
CURRENT APPLICATION NUMBER: US/10/027,805
FILING DATE: 21-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/602,359
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: HALE, LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
TELEFAX: 619-678-5099
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 249 AMINO ACIDS
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-10-027-805-39

Query Match 70.8%; Score 901.5; DB 9; Length 249;
Best Local Similarity 71.5%; Pred. No. 4.3e-83;
Matches 176; Conservative 31; Mismatches 38; Indels 1; Gaps 1;
QY 1 LRLKKEEINLVLSGGAAGIAHIGVLAINELGIRVPLSGVSAAGIYVFAASGSP 60
DB 5 LRLKKEEINLVLSGGAAGIAHIGVLAINELGIRVPLSGVSAAGIYVFAASGSP 60
QY 61 GMSFLKRVNMLKLFKRPPLKGLIGWEKAIPTLEEVLPYRRIEKLPIPTICADLVSG 120
DB 61 GMSFLKRVNMLKLFKRPPLKGLIGWEKAIPTLEEVLPYRRIEKLPIPTICADLVSG 120
QY 121 RALYLSGSLIPALGSCAIPGIFEPVEYKYLVDGIVNNLPVEPFOESGIPYCVDV 180
DB 121 RALYLSGSLIPALGSCAIPGIFEPVEYKYLVDGIVNNLPVEPFOESGIPYCVDV 180
QY 181 LPIPEPKDKNLIHLILSFFLAVNSSEKKEPCDVIYVPELBEFTPLDVRKADQIMER 240
DB 181 LPIPEPKDKNLIHLILSFFLAVNSSEKKEPCDVIYVPELBEFTPLDVRKADQIMER 240
QY 241 GYIKAL 246
DB 244 GDMRAL 249

RESULT 6
 US-10-027-804-39
 ; Sequence 39, Application US/10027804
 ; Publication No. U620030054530A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ROBERTSON, Daniel E.
 ; MURPHY, Dennis
 ; REID, John
 ; MAFFIA, Anthony
 ; LINK, Steven
 ; SWANSON, Ronald V.
 ; WARREN, Patrick V.
 ; KOSMOTKA, Anna
 ; TITLE OF INVENTION: ESTERASES
 ; NUMBER OF SEQUENCES: 42
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: FISH & RICHARDSON P.C.
 ; STREET: 4225 EXECUTIVE SQUARE, STE 1400
 ; CITY: LA JOLLA
 ; STATE: CALIFORNIA
 ; COUNTRY: USA
 ; ZIP: 92037
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 INCH DISKETTE
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: WORD PERFECT 6.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/027,804
 ; FILING DATE: 21-Dec-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/602,359
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: HALL, LISA A.
 ; REGISTRATION NUMBER: 38,347
 ; REFERENCE/DOCKET NUMBER: 09010/010001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 619-678-5070
 ; TELEFAX: 619-678-5099
 ; INFORMATION FOR SEQ ID NO: 39:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 249 AMINO ACIDS
 ; TYPE: AMINO ACID
 ; TOPOLOGY: LINEAR
 ; MOLECULE TYPE: PROTEIN
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 39:
 US-10-027-804-39
 Query Match 70.8%; Score 901.5; DB 9; Length 249;
 Best Local Similarity 71.5%; Pred. No. 4.3e-83;
 Matches 176; Conservative 31; Mismatches 38; Indels 1; Gaps 1;
 QY 1 LRLKPEEINLVSGGAAGIAHIGVLKAINELGIRVRLSGVSAGAIUSVVFASGSP 60
 DB 5 LRLKPEEINLVSGGAAGIAHIGVLKAINELGIRVRLSGVSAGAIUSVVFASGSP 64
 QY 61 GMSLLKRVNWLKFKPKPLKGLIGWEKAIKFLVEVLPYRRIEKLPIYICATDLYSG 120
 DB 65 EMLKLLKRVNWLKFKPKPKPLKGLIGWEKAIKFLVEVLPYRRIEKLPIYICATDLYSG 124
 QY 121 RALYLSGSLIPALIGSCAIPGIFEPVEYKNNLVLDGGIVNNLPVPEPQESGIPVCDV 180
 DB 125 KALYFGGDLIPVLLGSKSIFGIFEPVEYENFLVDGIVNNLPVPEPQESGIPVCDV 184
 QY 181 LPIEPKQDKNHLILRSFFLAVERNSKEKFCEDLVIVPELEETPTLDVKAQIMER 240
 DB 185 LPIEPKQDKNHLILRSFFLAVERNSKEKFCEDLVIVPELEETPTLDVKAQIMER 243
 QY 241 GYNAL 246
 ; : : :

DB 244 GDMRAL 249
 RESULT 7
 US-10-147-026-10
 ; Sequence 10, Application US/10147026
 ; Publication No. US20030003538A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Dietrich, Paul Shartzler
 ; APPLICANT: Huang, Chiao-Chain
 ; APPLICANT: Johnson, Carl D.
 ; APPLICANT: Sangameswaran, Lakshim
 ; TITLE OF INVENTION: Neuropathic Pain Genes; Compositions
 ; FILE REFERENCE: ROCH-006
 ; CURRENT APPLICATION NUMBER: US/10/147,026
 ; CURRENT FILING DATE: 2002-05-15
 ; PRIOR APPLICATION NUMBER: 60/155,702
 ; PRIOR FILING DATE: 1999-09-23
 ; PRIOR APPLICATION NUMBER: 60/189,931
 ; PRIOR FILING DATE: 2000-04-04
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 10
 ; LENGTH: 425
 ; TYPE: PRT
 ; ORGANISM: rattus sp.
 US-10-147-026-10
 Query Match 11.0%; Score 139.5; DB 9; Length 425;
 Best Local Similarity 23.2%; Pred. No. 1e-05;
 Matches 64; Conservative 41; Mismatches 104; Indels 67; Gaps 12;
 QY 10 NLVLSGGAAGIAHIGVLKAINELGIRV----RALSGVSAGAIUSVVFASG----YSPEG 61
 DB 9 SLSFAGCGFLGFYHIGATLCLSERAPHILREARFFGCSAGALHAVTFVCSLPDLHIMEI 68
 QY 62 MFSLLKRVNWLKFKPKPLKGLIGWEKAIKFLVEVLPYRRIEKLPIYICATDLYSG 120
 DB 69 LMDLVKARSNRNIGTLHP-----FFNINKCVRDQGLQETLPDNVHIOIISGVYISLTVRSOG 134
 QY 121 RALYLSGSLIPALIGSCAIPGIFEPVEYKNNLVLDGGIVNNLP-----164
 DB 125 ENLVSEFHSKDEVDALVCSFCIFLPSGLIPP-SFRCERYVDGSDNVFVLDKTTIT 183
 QY 165 VSPFQESGIPVCDVLPVLE-PEKDKNI-----LHILRSFFLAVERNSKEKFC 215
 DB 184 VSPFY--GEHDICPKVASTNFIQVNTNLSLCTGNLHLLTRALF-----227
 QY 216 DLVIVPELEETPTLDVKAQIMERGIYKALEVLSR 251
 DB 228 -----PSDVKVMGELCFQGYLDARFLEE 251
 RESULT 8
 US-09-765-205-40
 ; Sequence 40, Application US/09765205
 ; Patent No. US20020034800A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Li
 ; TITLE OF INVENTION: BONE MARROW SECRETED PROTEINS AND POLYNUCLEOTIDES
 ; FILE REFERENCE: 1458.004/200130.449
 ; CURRENT APPLICATION NUMBER: US/09/765,205
 ; CURRENT FILING DATE: 2001-01-17
 ; PRIOR APPLICATION NUMBER: US/09/212,440
 ; PRIOR FILING DATE: 1998-12-16
 ; NUMBER OF SEQ ID NOS: 46
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 40
 ; LENGTH: 504
 ; TYPE: PRT
 ; ORGANISM: human
 US-09-765-205-40

Query Match 8.6%; Score 109.5; DB 10; Length 504;
 Best Local Similarity 26.0%; Pred. No. 0.014;
 Matches 69; Conservative 32; Mismatches 115; Indels 49; Gaps 12;

10 NLVSGGAAKGIHIGVLTAKINE-----IGIRVAAALSGVAGAIIVSVFYASGYSPEGMFSL 65
 9 NISPGCGFLGYVYVGVASCSREHAPLVANAHITGASGALTALATVGVCGAGAGAK 68
 66 LKRVNMLKFKRKPLKGLIGMEKALP-EEVVLPRRIEKLPIPTICADLYSGRLY 124
 69 FLEVKARKEPFLGPHSPFLVLTISFLKTLPADSHHAGSGRLGISTFVSGENVYI 128
 125 L-----SEGLIPALLGSCAIP---GIEP-----VEYKNYLLVDGIVNNLP----- 164
 129 IGHFMSKDELIONAVCSGFIPYCGILPRLQGVY-----VDGISTDLPIYELKNIT 183
 165 VERPQESGIPYVCVDVLFIEPEKDKYILHILRSFPLAVRSNKKRKEFCDLVIVPELE 224
 184 VSPF--SGESDLC-----PQDSSTIHR--LRVNTSIQFNLNLYLSKALFPPB-- 230
 225 EFTPLDVRKADQIMERGVIKALEVL 249
 231 ---PLVLR---EMCKQGYRDLGLRFL 249

RESULT 9
 US-10-034-937-14

Sequence 14, Application US/10034937
 Publication No. US20030097684A1
 GENERAL INFORMATION:
 APPLICANT: Carri, Brian
 APPLICANT: Rosen, Barbara A.
 APPLICANT: Bermudez, Ericka R.
 APPLICANT: Ness, Jon E.
 APPLICANT: Maxygen, Inc.
 TITLE OF INVENTION: Lipid Acyl Hydrolases and Variants Thereof
 FILE REFERENCE: 18097A-032010US
 CURRENT APPLICATION NUMBER: US/10/034,937
 PRIOR FILING DATE: 2002-03-29
 PRIOR APPLICATION NUMBER: US 60/260,477
 NUMBER OF SEQ ID NOS: 47
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 14
 LENGTH: 390
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: clone PIP-10
 OTHER INFORMATION: Improved pentin lipid acyl hydrolase
 US-10-034-937-14

Query Match 6.8%; Score 86.5; DB 9; Length 390;
 Best Local Similarity 22.7%; Pred. No. 2.1;
 Matches 53; Conservative 30; Mismatches 79; Indels 71; Gaps 11;

11 LVLSGGAAGIAGHIGVLTAKIN-----ELGIRVAAALSGVAGAIIVSV----- 51
 21 LAIDGGGIRGIIIPGILTKQLATLQRMDSARLAAYFDVAVAGSITGIIITALLTPADPQN 80
 52 ---FYASG---YSPGMSLAKRVNMLKFKRKPLKGLIGMEKALPFEVLPYRR 102
 81 KDRPLAAGEIIDFIEHGPSEIFNKSPTACSSPGIFGCKYDG-----KYIQETIISQKL 132
 103 IEKL-----ELPTVICADLYSGALYISGSL--IPAL-----LQSCAIPGIF 144
 133 NETLDDQTTWVIVPSF---DILKLPITFSFPLKEVPELAVNKLSDVCWGTSAAPYIVF 186
 145 EEVEYKN-----YLVDGGIVNNLPVEBPQESGIPYVCVDVLFIEPEKDKYIL 193
 189 PPIYFRAGDTFENLVDAIINDIFA-----FVALSEVLQOEYKYN-KEIL 232

RESULT 10
 US-10-160-758-11

Sequence 11, Application US/10160758
 Publication No. US20030036076A1
 GENERAL INFORMATION:
 APPLICANT: EXELIXIS, INC.
 TITLE OF INVENTION: CADS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
 FILE REFERENCE: EX02-083C
 CURRENT APPLICATION NUMBER: US/10/160,758
 PRIOR FILING DATE: 2002-06-03
 PRIOR APPLICATION NUMBER: US 60/296,076
 PRIOR FILING DATE: 2001-06-05
 PRIOR APPLICATION NUMBER: US 60/328,605
 PRIOR FILING DATE: 2001-10-10
 PRIOR APPLICATION NUMBER: US 60/357,253
 PRIOR FILING DATE: 2002-02-15
 NUMBER OF SEQ ID NOS: 16
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 11
 LENGTH: 3354
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-160-758-11

Query Match 6.7%; Score 85.5; DB 9; Length 3354;
 Best Local Similarity 21.3%; Pred. No. 54;
 Matches 53; Conservative 36; Mismatches 75; Indels 85; Gaps 11;

22 AHIGVLR-----AINEIGIRVRA-----LSGVSAGAIIVSVFYASGYSPEGMFSL 65
 565 AYVALRENEPSVTQL-VLRAATDESPNNQITTSYISASAFSGYSDISYSGYIVSY 623
 66 LKRVNMLKFKRKPLKGLIGMEKALP-EEVVLPRRIEKLPIPTICADLYSGRLY 124
 624 SRPLDYEQISNGILYFTVAMDMGNPPLNSTVPTIEVDENDNPFTSKRPAFVSIVY-- 681
 100 YRIEKLPIPTICADLYSGR-----ALYSEGL--IPALLGSCAIPGIFPEVPEYK 150
 682 -ENIMAGATVLPFNATDLDRSREYQESIIYELSGSTQFRINASGEITTSLLDRETKS 740
 151 NYLL-----VDGIVNNLPVEBPQESGIPYVCVDVLFIEPEKDKYILHILRSFPL 186
 741 EYLIYRAVDGSGVGN-----QKTGLATVNTITLIDINDNHPYKADPYIINLVEMTPPD 794
 187 KQINKLHI 195
 795 SDVTTVVAV 803

RESULT 11
 US-10-160-758-12

Sequence 12, Application US/10160758
 Publication No. US20030036076A1
 GENERAL INFORMATION:
 APPLICANT: EXELIXIS, INC.
 TITLE OF INVENTION: CADS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
 FILE REFERENCE: EX02-083C
 CURRENT APPLICATION NUMBER: US/10/160,758
 PRIOR FILING DATE: 2002-06-03
 PRIOR APPLICATION NUMBER: US 60/296,076
 PRIOR FILING DATE: 2001-06-05
 PRIOR APPLICATION NUMBER: US 60/328,605
 PRIOR FILING DATE: 2001-10-10
 PRIOR APPLICATION NUMBER: US 60/357,253
 PRIOR FILING DATE: 2002-02-15
 NUMBER OF SEQ ID NOS: 16
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 12
 LENGTH: 3354
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-160-758-12

Query Match 6.7%; Score 85.5; DB 9; Length 3354;
 Best Local Similarity 21.3%; Pred. No. 54;
 Matches 53; Conservative 36; Mismatches 75; Indels 85; Gaps 11;
 QY 22 AHGVLK-----AINEIGIRVRA-----LSGSAGAIVSFYASGYSPGMPSL 65
 DB 565 AYVGAURENPSVTQL-VRLRATDESDPPNQITYSIVSASAFSGYFDISLYGYGVISV 623
 QY 66 LKRVN-----LKLFPKPPKGLI-----GWKRAIRFLREVL 99
 DB 624 SRPLDVEQISNGLIYLTWAMDAGNPPLNSTVPTIEVDENDNPPTFSKPAYFVSV-- 681
 QY 100 YRIEKLPTTICATLYSGR-----ALYLSGSL--IPALLGSCAIPGIFPVEYK 150
 DB 682 -ENIMAGATVFLNLDLRSEYQESIIYLSGEGTQFRINARSGEITTTSLDRETWS 740
 QY 151 NYVL-----VDGGIVANLPVEPPQESGIPTCVDVLPPIE-----PE 186
 DB 741 EYLIIVRAVDGGVGHN-----QKTGIATVITLIDINDNHTWKDAPYVINVLTPTPD 794
 QY 187 KQIKNLIHI 195
 DB 795 SDVITVVAV 803

RESULT 12
 US-10-034-937-10
 ; Sequence 10, Application US/10034937
 ; Publication No. US20030097684A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Carr, Brian
 ; APPLICANT: Rosen, Barbara A.
 ; APPLICANT: Bermudez, Ericka R.
 ; APPLICANT: Ness, Jon E.
 ; APPLICANT: Maxygen, Inc.
 ; TITLE OF INVENTION: Lipid Acyl Hydrolases and Variants Thereof
 ; FILE REFERENCE: 18097A-032010US
 ; CURRENT APPLICATION NUMBER: US/10/034,937
 ; PRIOR FILING DATE: 2002-03-29
 ; PRIOR APPLICATION NUMBER: US 60/260,477
 ; PRIOR FILING DATE: 2001-01-08
 ; NUMBER OF SEQ ID NOS: 47
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 10
 ; LENGTH: 390
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: clone PIP-6
 ; OTHER INFORMATION: improved pentin lipid acyl hydrolase
 US-10-034-937-10

Query Match 6.6%; Score 83.5; DB 9; Length 390;
 Best Local Similarity 20.9%; Pred. No. 4.1;
 Matches 48; Conservative 40; Mismatches 77; Indels 65; Gaps 12;
 QY 11 LVLSGANGIAHIGVKAIN-----ELGIRVRLSGVSGAGAIVSFYASGYSP 60
 DB 21 LAIDGGGIRGIIIPGVILKOLEATLQWDSARLAEPDVAVAGTSTGGIITAILTAP-DPQ 79
 QY 61 G-----NFSLLKRVNWL-----KLFK-----FKPLKGLIGWKAIRFLREVL 101
 DB 80 NKGRLVNAEIIIFIEHGPSIFENKSTACSLPGIFCPKYDG-----KYLQEIISQK 131
 QY 102 -----REKLEIPYICATDLYGRALYLSGSL--IPAL-----LGSCAIPGIFPV 147
 DB 132 LNETRLDQTTNVVIFSFQIKLRPTFTFKLEVPPELVNKLSDVCMGTSAAPIVPPY 191
 QY 148 EYKN-----YLLVDGGVNNLPVEPPQESGIPTCVDVLPPIEPEKDIKNIL 193
 DB 192 YPKHGDTFNLVDGAIADIPA-----PVALSEVLQEKYKN-KEIL 232

RESULT 13
 US-10-082-627A-25
 ; Sequence 25, Application US/10082627A
 ; Publication No. US20030104595A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Thym, Detlef
 ; APPLICANT: Knappe, Wolfgang-Reinhold
 ; APPLICANT: Shao, Zhixin
 ; APPLICANT: Schmuck, Rainer
 ; APPLICANT: Kratzsch, Peter
 ; APPLICANT: Bunk, Daniela
 ; TITLE OF INVENTION: New Forms of Soluble Pyrroloquinoline Quinone-Dependent Glucos
 ; FILE REFERENCE: RDID 0012CIPUS
 ; CURRENT APPLICATION NUMBER: US/10/082,627A
 ; CURRENT FILING DATE: 2001-10-29
 ; PRIOR APPLICATION NUMBER: EP 00123512.6
 ; PRIOR FILING DATE: 2000-10-27
 ; PRIOR APPLICATION NUMBER: EP 00127294.7
 ; PRIOR FILING DATE: 2000-12-19
 ; PRIOR APPLICATION NUMBER: US 09/710197
 ; PRIOR FILING DATE: 2000-11-09
 ; NUMBER OF SEQ ID NOS: 26
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 25
 ; LENGTH: 455
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 US-10-082-627A-25

Query Match 6.6%; Score 83.5; DB 9; Length 455;
 Best Local Similarity 22.1%; Pred. No. 5.2;
 Matches 52; Conservative 27; Mismatches 67; Indels 89; Gaps 11;
 QY 7 BEINLVLSGG-----AAKGIAHIGVLKAINELGIRVRLSGVSGAGAIVSFYAS 55
 DB 252 DEINLVKGGNGWPNVAGYKDSGYAYANYSAATNKSQIKLQNGIKVATGVPTKES 311
 QY 56 GYSPGMSLLKRVNWLKFKPKPLKGLIGWKAIRFLREVLPRYREKLEIPT----- 110
 DB 312 EWTGK-----NFPPLKTLTYQDTYND-----FTCGMA 343
 QY 111 YICATDLYSGRALYLSGSLIPALLGSCAIPGIFPVEYKYLVDG--GIVNNLPVEP 167
 DB 344 YICWPTVAPSSA-YVITG-----GKKAIPG-----WENTLLVPSLKRGVIPRIKLP 389
 QY 168.FQESGIPTCVDVLPPIEPEKDIKNILHILLRFFLAVRSNRSERKEFCDLVIVE 222
 DB 390 TYS-----TTLDDAIPM-----FKSNRYR-----DVIASPE 416

RESULT 14
 US-09-815-242-5379
 ; Sequence 5379, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848

;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 5379
;; LENGTH: 391
;; TYPE: PRF
;; ORGANISM: Staphylococcus aureus
US-09-815-242-5379

Query Match 6.5%; Score 83; DB 10; Length 391;
Best Local Similarity 23.2%; Pred. No. 4.7;
Matches 62; Conservative 36; Mismatches 93; Indels 76; Gaps 14;

QY 32 ELGIRVALSGVS-AGATVSVFYASGSPGMSLSLRVWMLKLFKPKPLKGLIGMEKA 90
DB 142 EAAVDLAKLTGAKPAGVICIEMNDGTWAKGD-----LQFKERKQLKMIT----- 188
QY 91 IRPLEVLPYRR-----IE---KLEIPYICATDLYSGALYLSBSLI----- 131
DB 189 ---IDDLIEYRKKEPEIEFKAKVMPTDFGTHTMGFKATYTDERIVLTKGAIQHEN 245
QY 132 PALSGCAIRGIF-----EPVEYKN-----YLVYG---GIVNNLPVEPPO 169
DB 246 VRLHSACLTGDI FHSQRCDGAGQLESMSKYNIEHGMIIYLPQEGIGIGLNLKRAYELI 305
QY 170 ESGIPYCVDPVPIPEPKDIK-----NIIHLIRSFPLAVRSNSEKKEKCDL 217
DB 306 EGGYDVTYTN-LALGFDDELDRDHYHIAQLKFFIEHINLS---NNPSFESGAKQY-GI 360
QY 218 VIVPELEEFPLDVKRADOIMERGYIK 244
DB 361 DIARIEVIVPEYVHNHD-YMETKKIK 386

RESULT 15

US-09-815-242-12285
; Sequence 12285, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haseibeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zvekind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931

;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 12285
;; LENGTH: 393
;; TYPE: PRF
;; ORGANISM: Staphylococcus aureus
US-09-815-242-12285

Query Match 6.5%; Score 83; DB 10; Length 393;
Best Local Similarity 23.2%; Pred. No. 4.7;
Matches 62; Conservative 36; Mismatches 93; Indels 76; Gaps 14;

QY 32 ELGIRVALSGVS-AGATVSVFYASGSPGMSLSLRVWMLKLFKPKPLKGLIGMEKA 90
DB 144 EAAVDLAKLTGAKPAGVICIEMNDGTWAKGD-----LQFKERKQLKMIT----- 190
QY 91 IRPLEVLPYRR-----IE---KLEIPYICATDLYSGALYLSBSLI----- 131
DB 191 ---IDDLIEYRKKEPEIEFKAKVMPTDFGTHTMGFKATYTDERIVLTKGAIQHEN 247
QY 132 PALSGCAIRGIF-----EPVEYKN-----YLVYG---GIVNNLPVEPPO 169
DB 248 VRLHSACLTGDI FHSQRCDGAGQLESMSKYNIEHGMIIYLPQEGIGIGLNLKRAYELI 307
QY 170 ESGIPYCVDPVPIPEPKDIK-----NIIHLIRSFPLAVRSNSEKKEKCDL 217
DB 308 EGGYDVTYTN-LALGFDDELDRDHYHIAQLKFFIEHINLS---NNPSFESGAKQY-GI 362
QY 218 VIVPELEEFPLDVKRADOIMERGYIK 244
DB 363 DIARIEVIVPEYVHNHD-YMETKKIK 386

Search completed: June 19, 2003, 15:13:00
Job time: 183 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 19, 2003, 14:22:25 ; Search time 21 Seconds

(without alignments)
495.741 Million cell updates/sec

Title: US-09-903-410-36

Perfect score: 1273
Sequence: 1 LRLRRFEINILVSGGAAG.....RKADQIMRGYKALEVLSR 251

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt 40.1*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	268.5	21.1	260	1	YLRK_BACSU
2	257.5	20.2	314	1	YCHK_BCOLI
3	242	19.0	1679	1	YMF9_YEAST
4	237.5	18.7	192	1	YCHK_SHIFL
5	235	18.5	583	1	YPE5_MYCTU
6	234	18.4	360	1	YAE3_MYCTU
7	223.5	17.6	291	1	YORH_BACSU
8	217.5	17.1	1425	1	SWS_DROME
9	208.5	16.4	1371	1	YOL4_CAEEL
10	189.5	14.9	880	1	YU7_CAEEL
11	132.5	10.4	910	1	YK69_YEAST
12	124	9.7	621	1	YPO7_COEEL
13	121	9.5	253	1	G82_HUMAN
14	100.5	7.9	357	1	YJUT_BCOLI
15	98.5	7.7	435	1	FIXC_AZOCA
16	98.5	7.7	1073	1	YCP4_YEAST
17	94.5	7.4	769	1	HCPA_YEAST
18	92	7.2	531	1	FXN1_SCHRO
19	92	7.2	786	1	CGZP_HUMAN
20	92	7.2	1102	1	YE20_METUA
21	91	7.1	459	1	PURL_THEVO
22	90	7.1	469	1	SYEL_THEMA
23	89.5	7.0	588	1	CALI_HUMAN
24	87.5	6.9	397	1	HEM1_PYRAE
25	86.5	6.8	282	1	DH11_MOUSE
26	86.5	6.8	4427	1	PKSL_BACSU
27	86	6.8	456	1	PROL_YEAST
28	85.5	6.7	1161	1	DP3A_AQUAE
29	85	6.7	3354	1	CADN_HUMAN
30	84	6.6	255	1	UPPS_METTH
31	83.5	6.6	431	1	SYD_METKA
32	83.5	6.6	564	1	TP6B_PYRAB
33	83.5	6.6	773	1	SYFB_CAMDE

34	83	6.5	1884	1	RRE5_HUMAN	Q14690 homo sapien
35	82.5	6.5	292	1	DH11_HUMAN	P28845 homo sapien
36	82.5	6.5	435	1	FIXC_BRAJA	P10331 bradyrhizob
37	82.5	6.5	564	1	TP6B_PYRHO	O74020 pyrococcus
38	82.5	6.5	698	1	ALT_PYP2	Q38424 bacterioph
39	82	6.4	242	1	TRUA_HBLPU	Q92K00 helicobacte
40	82	6.4	496	1	SRM_MOUSE	O62270 mus musculu
41	82	6.4	948	1	SECA_SYNP7	O55357 synchococc
42	81.5	6.4	318	1	TEG_BACME	P30420 bacillus me
43	81.5	6.4	319	1	KGF_THEMA	O9WY32 thermotoga
44	81	6.4	729	1	SYG_MOUSE	Q9CZD3 mus musculu
45	81	6.4	759	1	NAH2_SCHPO	Q14123 schizosacch

ALIGNMENTS

RESULT 1	YLRK_BACSU	STANDARD;	PRT;	260 AA.
ID	YLRK_BACSU			
AC	O34731;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, last annotation update)			
DE	Hypothetical protein ylrk.			
GN	YLRK.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=98044033; PubMed=9384377;			
RA	Kunze F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,			
RA	Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,			
RA	Borries R., Bouteiller L., Brous S., Braun M., Brignell S.C., Bron S.,			
RA	Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,			
RA	Choi S.K., Codani J.J., Comerion I.F., Cummings N.J., Daniel R.A.,			
RA	Dentzov F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,			
RA	Bautan K.D., Brington J., Fabret C., Ferrati E., Foulger D.,			
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,			
RA	Ghm S.Y., Glaser P., Goffeau A., Golshety E.J., Grandi G.,			
RA	Guillepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,			
RA	Hilbert H., Holtschapel S., Hosono S., Hulio M.F., Itaya M., Jones L.,			
RA	Joris B., Karamata D., Kasahara Y., Kjaer-Blanchard M., Klein C.,			
RA	Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,			
RA	Kurita K., Lapidos A., Lardinois S., Lauber J., Lazarevic V.,			
RA	Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue V.,			
RA	Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,			
RA	Moore D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,			
RA	Parro V., Pohl T.M., Portetelle D., Portwollik S., Prescott A.M.,			
RA	Prescan E., Fujic P., Fumelle B., Rapoport G., Rey M., Reynolds S.,			
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,			
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scotlone F.,			
RA	Sekiguchi J., Sekouci B., Aker S.J., Serro P., Shin B.S., Soldo B.,			
RA	Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,			
RA	Takuchi M., Tanakoshi A., Tanaka T., Terstap P., Tognoni A.,			
RA	Tosato V., Uchiyama S., Vandenbol M., Vanter F., Vassarotti A.,			
RA	Vitari A., Wandut R., Wedler E., Wedler H., Weitzenecker T.,			
RA	Winters E., Wipat A., Yamamoto H., Yamane K., Yamamoto K., Yata K.,			
RA	Yoshida K., Yoshikawa H.F., Zumestein E., Yoshikawa H., Danchin A.,			
RT	"The complete genome sequence of the Gram-positive bacterium Bacillus			
RT	subtilis."			
RL	Nature 390:249-256(1997).			
CC	-1- SIMILARITY: BELONGS TO THE UPF0028 (SWS) FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			

DB 238 AVTAPATATITMTSIVLETKRNMADPPDILOVCCPISITLDPHNAIANGOL 237
 QY 244 KALETSE 251
 DB 298 AVERKXDE 305

RESULT 3
 YMF9_YEAST STANDARD; PRT; 1679 AA.
 ID YMF9_YEAST
 AC Q04958;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 187.1 kDa protein in OG4-CNA2 intergenic region.
 GN YML059C OR YMW958.03C
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Buxarjota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=528bc / AB972;
 RA Devlin K., Churcher C., Barrall B.G., Rajandream M.A.;
 RL Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE UPF0028 (SMS) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).

DR EMBL; Z46729; CAAB6716.1; -
 DR SGD; S0004524; YML059C.
 DR InterPro; IPR002641; Patatin.
 DR InterPro; IPR001423; UPF0028.
 DR InterPro; IPR000595; CNMP binding.
 DR Pfam; PF00027; CNMP binding; 2.
 DR Pfam; PF01734; Patatin; 1.
 DR SMART; SM00100; CNMP; 1.
 DR PROSITE; PS50042; CNMP_BINDING_3; 2.
 DR PROSITE; PS01237; UPF0028; 1.
 DR Hypothetical protein; Transmembrane.
 FT TRANSMEM 50 70 POTENTIAL.
 FT TRANSMEM 104 124
 SEQUENCE 1679 AA; 187132 MW; 550PFCDAACAF8E25 CRC64;

Query Match 19.0%; Score 242; DB 1; Length 1679;
 Best Local Similarity 28.5%; Pred. No. 1.2e-13;
 Matches 79; Conservative 42; Mismatches 120; Indels 36; Gaps 6;

QY 1 LRLAKF---EIMVLSGSAKGIAGIIVKALINEIGIRVRLSGVSAGATVSFYVAGSY 57
 DB 1360 LRLAKLISGALIGVAGGARGISHLGIVQAIIEGQIFVDVIGSTISGVGLVAKDY 1419

QY 58 SPBGMSFLK---RVN--WLKLFKRPPLKGLIGMEKAIPLSEVLPYRIEKLPTPT 111
 DB 1420 DLVPIYGRVKKAFGRISIRMLTDLTWPTVSYTGHFNRGIVWTPGDTIEIDPFWIOY 1479

QY 112 ICATDLVSGALYLSGSLIPALISGCAIPGIFEPVEYKAYLLVDGIVNNLPVPEPQES 171
 DB 1480 CNGSTNINIDSVQEHSPFYARIRYIRASMSLAGLPLLENGSVLLDGGVVDLPTENBAR 1539

QY 172 GIPTV-CVDV-----LPIPEKDIKNILHILIRFPFLAVR 205
 DB 1540 GCQTIPADVGSADDTPEYSGDSLNGFWIIFRRMNPFSSSHPIIDMAIQVRLGYAVSV 1599

QY 206 SNSEKREKCDIVIV-PELEEFPTLDVRKADQIMERG 241

DB 1600 MALEKAKTRGVVYVVRPIEBYATIDFSKFEIYHNG 1636

RESULT 4
 YCHK_SHIFL STANDARD; PRT; 192 AA.
 ID YCHK_SHIFL
 AC P37054;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 34.4 kDa protein in HNR-PUR0 intergenic region
 DB (Fragment).
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Shigella.
 OX NCBI_TaxID=623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Serotype 2A;
 RX MEDLINE=93023838; PubMed=1406252;
 RA Hromockyj A.E., Tucker S.C., Maurelli A.T.;
 RT "Temperature regulation of Shigella virulence: identification of the
 RT repressor gene vtr, an analogue of hns, and partial complementation
 RT by tyrosyl transfer RNA (tRNA^{Tyr}).";
 RL Mol. Microbiol. 6:2113-2124(1992).
 RN [2]
 RP CONCEPTUAL TRANSLATION.
 RA Ridd K.E.;
 RL Unpublished observations (FEB-1994).
 CC -1- SIMILARITY: BELONGS TO THE UPF0028 (SMS) FAMILY.
 CC -----
 CC BR INTRODUCED IN POSITION 178 TO PRODUCE THIS ORF.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).

DR EMBL; X66849; -; NOT_ANNOTATED_CDS.
 DR InterPro; IPR002641; Patatin.
 DR InterPro; IPR001423; UPF0028.
 DR Pfam; PF01734; Patatin; 1.
 DR PROSITE; PS01237; UPF0028; 1.
 DR Hypothetical protein.
 FT NON_TER 192
 SEQUENCE 192 AA; 20831 MW; 34AEFGA693CC917A CRC64;

Query Match 18.7%; Score 237.5; DB 1; Length 192;
 Best Local Similarity 33.7%; Pred. No. 2.1e-14;
 Matches 58; Conservative 33; Mismatches 64; Indels 17; Gaps 4;

QY 3 LRKEEINLVLSGSAKGIAGIIVKALINEIGIRVRLSGVSAGATVSFYVAGSY 62
 DB 14 MKKI-KALALSGSARGSHGIVNALKKVCEIEDIYAGSISIVGAAYAC-----DR 67

QY 63 FSLIKRVNLIKLFKPKPK-----GLIGMEKAIPLSEVLPYRIEKLPTPT 113
 DB 68 LSALE--DWTSPSYVELRLIMDLWSORGLRGRVNOYREIMPEIEHNCRRPAPV 125

QY 114 ATDYSGRALYLSGSLIPALISGCAIPGIFEPVEYKAYLLVDGIVNNLPV 165
 DB 126 ATNISTREIMFTGDLHLAIRASCSIPGLMAPVAHNYMTVVDGAVNPIPI 177

RESULT 5
 YP65_MYCTU STANDARD; PRT; 583 AA.
 ID YP65_MYCTU
 AC Q50733;
 DT 01-NOV-1997 (Rel. 35, Created)

01-NOV-1997 (Rel. 35, Last sequence update)
 15-JUN-2002 (Rel. 41, Last annotation update)
 Hypothetical protein RV2565
 RV2565 OR MT2641 OR MT1934.03C.
 OS Mycobacterium tuberculosis
 OC Bacteria; Actinobacteria; Actinobacteriia (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 NCBI_TaxID=1773;
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=H37Rv;
 RC MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sultston J.E., Taylor K., Whittam S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE UPF0028 (SWS) FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z77250; CAN01050.1; -
 CC EMBL; AE007098; AKA46954.1; -
 CC TIGR; MT2641; -
 CC TubercuList; RV2565; -
 CC InterPro; IPR002641; Patatin.
 CC InterPro; IPR001423; UPF0028.
 CC Pfam; PF00027; cAMP binding.
 CC Pfam; PF01734; Patatin; 1.
 CC SMART; SM00100; cAMP; 1.
 CC PROSITE; PS0042; cAMP BINDING 3, 1.
 CC PROSITE; PS01237; UPF0028; 1.
 CC Hypothetical protein; Complete proteome.
 FT CONFLICT 375 375 G -> D (IN REF. 2).
 SQ SEQUENCE 583 AA; 62123 MW; EC6EE062BAEC62F7 CRC64;
 Query Match 18.5%; Score 235; DB 1; Length 583;
 Best Local Similarity 24.7%; Pred. No. 1.4e-13;
 Matches 70; Conservative 51; Mismatches 106; Indels 56; Gaps 6;
 QY 9 INLVSGGAAGIAHIGVKAINEIGIRVRLSGVAGAIIVSVFYA----- 54
 DB 307 LGLVMAGGARGIAHFGVYQELTEAGVVIDRFGTSGAIAASAFAALMGADGADIAAARE 366
 QY 55 --SGVSPEGMESLLKRVNWLKFKPKLGLGWKAKRFLKLEVLPRRIKLEIPTYI 112
 DB 367 FIAGSDPLGDYII-----PISALTRGGRVRLVGVFGVNTLIEHLRGRFSS 412
 QY 113 CATDLYSGRAVLYSGSLIPALGSCAIPGIFPEVYKYNLLVDDGGINVNLVPEPF-QES 171

DB 413 VSADMITGDIHRRGSGVAVRASISIFGLPPVHNGEQLLVGGGLNNLPANVMCACT 472
 QY 172 GIFTVCYD-----VLP--IEPEKDKNLL-----HILERSFFLAURS 206
 DB 473 DGEVICVDLRTFTVPSKGFGLPIVTPGGLLRLLTCTDNALPLOTETLLRDLAANT 532
 QY 207 NSERKKEFCDLVITVPELEBFTPLDVRKADQIMERGVIKALEVL 249
 DB 533 ANLRELPRVAATIEPDVSKIGVLPKQIDAALAEAGRAAAL 575
 RESULT 6
 YAG3 MYCTU
 ID YAG3 MYCTU STANDARD; PRT; 360 AA.
 AC O53411;
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein RV1063c.
 GN RV1063C OR MT1093 OR MT017.16C.
 OS Mycobacterium tuberculosis
 OC Bacteria; Actinobacteria; Actinobacteriia (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 NCBI_TaxID=1773;
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=H37Rv;
 RC MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sultston J.E., Taylor K., Whittam S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE UPF0028 (SWS) FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AL021897; CAN17179.1; -
 CC EMBL; AE006990; AKA45347.1; -
 CC TIGR; MT1093; -
 CC TubercuList; RV1063c; -
 CC InterPro; IPR002641; Patatin.
 CC InterPro; IPR001423; UPF0028.
 CC Pfam; PF01734; Patatin; 1.
 CC PROSITE; PS01237; UPF0028; FALSE NEG.
 CC Hypothetical protein; Complete proteome.
 SQ SEQUENCE 360 AA; 37522 MW; F08C30CE6096D0F6 CRC64;
 Query Match 18.4%; Score 234; DB 1; Length 360;

CC or send an email to license@sib-sib.ch).
 DR EMBL; M98552; AAA28206.2; -
 DR EMBL; M98552; AAL11481.1; -
 DR WormPep; ZK370.4a; CE25687.
 DR WormPep; ZK370.4b; CE29641.
 DR InterPro; IPR002641; Patatin.
 DR InterPro; IPR001423; UPR0028.
 DR InterPro; IPR000595; cNMP binding.
 DR Pfam; PF00027; cNMP_binding; 3.
 DR Pfam; PF01734; Patatin; 1.
 DR SMART; SM00100; cNMP; 2.
 DR PROSITE; PS50042; cNMP_BINDING_3; 3.
 DR PROSITE; PS01237; UPR0028; 1.
 KM Hypothetical protein: Alternative splicing.
 FT VASPPIC 512 522 MISSING (IN ISOFORM A).
 SQ SEQUENCE 1371 AA; 154528 MW; DADP21707BD47782 CRC64;
 Query Match 16.4%; Score 208.5; DB 1; Length 1371;
 Best Local Similarity 27.1%; Pred. No. 1e-10;
 Matches 75; Conservative 39; Mismatches 110; Indels 53; Gaps 9;
 QY 9 INLVSGAAKGIHIGVKAINEIGIRVRLSGVSGAIVSVFVYAGSPSE----- 60
 DB 958 IGVVGGGAGAAHAGALRALIEKQIDMVGSTISGALFGLVAT--TPDIRAVGRMK 1015
 QY 61 -----GMSILKRVNMLKFKKPKLGLIGMEKAIKRLPELRYRIEKLIPYICAT 115
 DB 1016 ASFWNGMSL-----WRKLLDTVAHSMFTGAGNFESIKDLFEERLIEDLWISYFCIST 1070
 QY 116 DLVSGRALYSESGSLIPALGSCAIPGIFEPV--EYKNYLLVDGGINNLPVPEPQSG 173
 DB 1071 DISTSEKVRKRSPLMAYGRASMSLAGTIPICPDQGHLLDGGYVNNVPADVMKRLA 1130
 QY 174 P-TVCVDVLPIT-----PEKDKNLIHLISFPLAVS 206
 DB 1131 RCYIACDVGSIEETNLYDGSISGMMWVLLKRLNFTGTPPR-IINMEIOSRLIAYSVCR 1189
 QY 207 NSR--KRKEFCDLVIVPELEFPTPLDVKRAQDIEMRG 241
 DB 1190 QLEVVKKASYC-RYLRPIEPKKTIDPFKQEIHELGS 1225
 RESULT 10
 ID YVL7_CABEL STANDARD; PRT; 880 AA.
 AC Q21534; Q20023;
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein M10.7 in chromosome II.
 GN M10.7.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peioderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Thomas K., Baynes C.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE UPR0028 (SMS) FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.emb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 DR EMBL; Z49966; CAA90246.1; -
 DR EMBL; Z49968; CAA90246.1; JOINED.

DR EMBL; Z49968; CAA90246.1; -
 DR EMBL; Z49966; CAA90246.1; JOINED.
 DR WormPep; M10.7; CE03510.
 DR InterPro; IPR002641; Patatin.
 DR InterPro; IPR001423; UPR0028.
 DR InterPro; IPR000595; cNMP binding.
 DR Pfam; PF00027; cNMP binding; 1.
 DR Pfam; PF01734; Patatin; 1.
 DR SMART; SM00100; cNMP; 2.
 DR PROSITE; PS50042; cNMP_BINDING_3; 1.
 DR PROSITE; PS01237; UPR0028; 1.
 KM Hypothetical protein.
 SQ SEQUENCE 880 AA; 100053 MW; A94D9EFED3A7DEED CRC64;
 Query Match 14.9%; Score 189.5; DB 1; Length 880;
 Best Local Similarity 23.0%; Pred. No. 3e-09;
 Matches 65; Conservative 45; Mismatches 113; Indels 59; Gaps 8;
 QY 9 INLVSGAAKGIHIGVKAINEIGIRVRLSGVSGAIVSVFVYAGSPSE----- 60
 DB 551 IGVVGGGAGAAHAGALRALIEKQIDMVGSTISGALFGLVAT--TPDIRAVGRMK 608
 QY 61 -----GMSILKRVNMLKFKKPKLGLIGMEKAIKRLPELRYRIEKLIPYICAT 111
 DB 609 DFTDTLRNNILDVVDLTN-----PYGILTGKRPVLCYQRLNDVNIEDCVSF 660
 QY 112 ICATDLYSGRALYSESGSLIPALGSCAIPGIFEPV--EYKNYLLVDGGINNLPVPEPQ 169
 DB 661 CITDTLTSSSMHINHGIMPEVVSMSIAGYVPICDPDQGHLLDGGYVNNLPADIMR 720
 QY 170 ESGIPV-CVDV-----LLEPEKDKNLIHLISFPL 202
 DB 721 SLCANVAVIADGMSDDNTNKAQFSTGWCLEFKRWPFGEELIIVLNNVQRLAIY 780
 QY 203 AVRSNSE--KRKEFCDLVIVPELEFPTPLDVKRAQDIEMRG 242
 DB 781 CCVQOMEIVKNAQCYVYLP-IRSGIFDPSKPDQAOIGY 821
 RESULT 11
 ID YK69_YEAST STANDARD; PRT; 910 AA.
 AC P36165;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 102.7 kDa protein in PRP16-SR640 intergenic region.
 GN YK089C OR YK0409.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=94262327; PubMed=8203164;
 RA Garcia-Cantalejo J., Baladron V., Estreban P.F., Santos M.A., Bou G.,
 RA Remacha M.A., Revuelta J.L., Ballesca J.P.G., Jimenez A., del Rey F.;
 RT "The complete sequence of an 18,002 bp segment of Saccharomyces
 RT cerevisiae chromosome XI contains the HBS1, MRP-120 and PRP16 genes,
 RT and six new open reading frames".
 RL Yeast 10:231-245 (1994).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: TO YEAST YK0313C AND S.POMBE SPC1450.16C.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.emb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 DR EMBL; Z27116; CAA81640.1; -

DR EMBL: Z28314; CA882168.1; --
 DR PIR: S38167; S38167.
 DR PIR: S39130; S39130.
 DR SCD: S0001797; YK0089C.
 DR InterPro: IPR002641; Patatin.
 DR Pfam: PF01734; Patatin; 1.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 282 302 POTENTIAL.
 FT TRANSMEM 310 326 POTENTIAL.
 FT TRANSMEM 426 442 POTENTIAL.
 SQ SEQUENCE 910 AA; 102716 MW; 1CF03C4AGE64B9C CRC64;
 Query Match 10.4%; Score 132.5; DB 1; Length 910;
 Best Local Similarity 30.4%; Pred. No. 0.00045;
 Matches 62; Conservative 22; Mismatches 75; Indels 45; Gaps 9;
 QY 11 LVLSGGAAGIAHIGVKAINELG--IRVRLSGVSAGAIYVVFVAGSYSGEGMFLSLL--- 66
 Db 282 LVLSGGTGFGLHIGVGLTFLDLPVRLSSGSAIYVSI--LSVHHKEIPVLLNHIL 340
 QY 67 -KRVNWLKLFKPKPLKGLI-----GW-----EKAIRFLEVL--PYRIEK 105
 Db 341 DKFPIFKDKKSESENLKISRFKNGTWFEDNKHVNTWIFGLGTLTFRAYNRGK 400
 QY 106 LEIPYVICATLYSGRAL--YLSGSLI--PALLGSCAIFGIF--RPVEYK----- 150
 Db 401 ILNITVSPASLPEQPRLLNLTAPNVLIWSAVCASCSLPGIIPSSPLYEKDKPKRKPW 460
 QY 151 ----NYLVGSGIVNNLVEPFQ 170
 Db 461 TGSSSVKFDGSDVNDLPISLSE 484
 RESULT 12
 ID YPD7 CABEL STANDARD; PRT; 621 AA.
 AC Q11136;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein C05D11.7 in chromosome III.
 GN C05D11.7.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2.
 RA Du Z.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP REVISIONS.
 RA Waterston R.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO HUMAN GS2 PROTEIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL: U00048; AE53832.2;
 DR Wormpep; C05D11.7; CE29663.
 DR InterPro; IPR002641; Patatin.
 DR Pfam; PF01734; Patatin; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 621 AA; 69363 MW; 93BE89A868F684EC CRC64;
 Query Match 9.7%; Score 124; DB 1; Length 621;

Best Local Similarity 27.6%; Pred. No. 0.0016;
 Matches 58; Conservative 28; Mismatches 72; Indels 52; Gaps 11;
 QY 7 BEINLVLSGGAAGIAHIGVKAINELG--IRVRLSGVSAGAIYVVFVAGSYSGEGMFLSLL--- 63
 Db 9 ELMNLSFGCGPLCVVHAGVAAAIKEAYAPELLQKILGASAGSIYVACGLITGVCIASHATS 68
 QY 64 SLKRVNWLKLFKPKP-----PLKGLIGWEKAIRFLEVLVPRRIEKLIEITYICATDLY 118
 Db 69 TLKAVSQASRTFGPLHPEFNLGIVRDE-----LEVILP-----EN--AVEMC 111
 QY 119 SGRAL-----YLSGSLIPALLGSCAIP-----GIFEPVEYKXVLLVUGGI 159
 Db 112 TGRVLISLTWSDHENVVIDEYRSNADLIDAIMCSFIPLYCGI--TPPKFRGVQYIDGGV 170
 QY 160 VNNLVEPFQESGIPPTVCDVLPFIEPKDI 189
 Db 171 SDNQPI--YDEH-----TIVTSFPFSGESDI 193
 RESULT 13
 ID GS2 HUMAN STANDARD; PRT; 253 AA.
 AC P41247;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE GS2 protein (DXS1283E).
 GN GS2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain.
 RX MEDLINE=95104848; PubMed=7806223;
 RA Lee W.-C., Salido E., Yen P.H.;
 RT "Isolation of a new gene GS2 (DXS1283E) from a CpG island between STS
 RT and KALI on Xp22.3.";
 RL Genomics 22:372-376(1994).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES EXAMINED, INCLUDING
 CC HEART, BRAIN, PLACENTA, LUNG, LIVER, MUSCLE, KIDNEY, PANCREAS AND
 CC SPLEEN.
 CC -1- SIMILARITY: TO C.ELEGANS C05D11.7.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL: U03886; AA16491.1;
 DR EMBL: U08893; AA17838.1;
 DR EMBL: U08888; AA17838.1; JOINED.
 DR EMBL: U08889; AA17838.1; JOINED.
 DR EMBL: U08890; AA17838.1; JOINED.
 DR EMBL: U08891; AA17838.1; JOINED.
 DR EMBL: U08892; AA17838.1; JOINED.
 DR MIM; 300102;
 DR InterPro; IPR002641; Patatin.
 DR Pfam; PF01734; Patatin; 1.
 SQ SEQUENCE 253 AA; 27980 MW; 6FFA94F460BA03AC CRC64;
 Query Match 9.5%; Score 121; DB 1; Length 253;
 Best Local Similarity 23.9%; Pred. No. 0.001;
 Matches 64; Conservative 47; Mismatches 109; Indels 48; Gaps 14;
 QY 7 BEINLVLSGGAAGIAHIGVKAINELGIR---VRALSGVSAGAIYVVFVAGSYSGEGM 62
 Db 2 KHINLSFAACGFLGIYHUGAASALCRHGKLVKDKAPAGASAGSILVASVLLT--APE-- 57


```

QY 63 FSLKRVNMLKLFKEPEPK-----GLI--GWEKAIKRP---LEEVYPRRIEKLPIFYIC 113
DB 58 --KIEECNOR-TYKFAEBIRROSFGAVTFPGYDFMARLRSQSGSTILPSAHLAONRLHVS 114
QY 114 ATDLYSGR---ALVYLSGSLIPALIGSCAIP---GIFEPYKXNYLLVDGIVNNLPVE 166
DB 115 ITAKATRENNHVLVTFSSREDLIKVLASSFVPIYAGL-KLVEYKQKQKVDGGLTNALPIL 173
QY 167 PFOESGIPYCVVULPIEBEKI---KNILHILRSFPLAVRSSEKREKFCOLVIVPE 222
DB 174 PVKRT-----VIFSPSGRLDISPODKQGLDLYV-----NIAKQDIMSILANDVR 218
QY 223 LEE-FTPLDVRKADQIMERGYIKALEVL 249
DB 219 LNOALFPPSKRKMEISLYCGFDDTVKFL 246

```

RESULT 14

```

YJUU_ECOLI STANDARD; PRT; 357 AA.
ID YJUU_ECOLI
AC P39407;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypochemical protein yj3U.
GN YJUU OR B4377.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes";
RL Nucleic Acids Res. 23:2105-2119(1995).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U14003; AAA97273.1; -.
DR EMBL; AB000508; AAC77330.1; -.
DR Ecogene; BG12597; yj3U.
DR InterPro; IPR002641; Pfatatin.
DR Pfam; PF01734; Pfatatin; 1.
KW Hypochemical protein; Complete proteome.
SQ SEQUENCE 357 AA; 39831 MW; D7A57185B93CCEBA CRC64;

```

Query Match 7.9%; Score 100.5; DB 1; Length 357;

Best Local Similarity 26.5%; Pred. No. 0.11; Indels 53; Gaps 12;

```

Matches 56; Conservative 30; Mismatches 72; Indels 53; Gaps 12;
QY 1 LRLRKEF--EINVLVSGAAGKIAHIGVAKINEIGIRVA-----LSGVAGA-IVS 50
DB 15 LSLRPFQPGRIALVCCGGGGRGIFTAGVILDEF-----MRAQPNFDDYLTGTSAGQNTLS 68
QY 51 VEVAS--GSPPEGNSLKLKRVNMLKLFKEPKLK-----GLIGWEKAIKRFLEEVLPYRRI 103
DB 69 AFTCNQPGYARKYI-----MRTYTRREFPDLRFRFGSNTLIDLMVYEARSGMP----- 118
QY 104 EKEIIPF-----YCA--TLYSGRALVYLSGSLIPALIGSCAIPGIFE---PV 147
DB 119 --LQMDTPARLFDGSKSFYWCACRODYAPNYFLPTKQNMMLDIVIRASSAIPIGFYRSQVSL 176.

```

```

QY 148 EYKNVLLVDGIGIVNNLPVEPFOESGIPYCV 178
DB 177 EGINVL--DGGISDAIVKKAARCAKATLVV 205

```

RESULT 15

```

FIXC_AZOCA STANDARD; PRT; 435 AA.
ID FIXC_AZOCA
AC P26484;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE FIXC protein.
GN FIXC.
OS Azorhizobium caulinodans.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Hyphomicrobium group; Azorhizobium.
OX NCBI_TaxID=7;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORS571;
RX MEDLINE=91203829; PubMed=1850088;
RA Arigoni F., Kaminski P.A., Hennecke H., Elmerich C.;
RT "Nucleotide sequence of the fixABC region of Azorhizobium caulinodans
RT ORS571: similarity of the fixB product with eukaryotic flavoproteins,
RT characterization of fixX, and identification of nifW.";
RL Mol. Gen. Genet. 225:514-520(1991).
CC
CC -1- FUNCTION: COULD BE REQUIRED FOR THE FORMATION OF A FUNCTIONAL
CC NITROGENASE FE PROTEIN. PROBABLY ACCEPTS ELECTRONS FROM FIXA/FIXB
CC AND REDUCES A QUINONE.
CC -1- COFACTOR: FAD (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE ETF-OO / FIXC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X55450; CAA39093.1; -.
DR EMBL; X55450; CAA39093.1; -.
DR PIR; S14072; S14072.
KW Nitrogen fixation; Oxidoreductase; FAD; Flavoprotein.
KW NF_BIND
FT SEQUENCE 435 AA; 48578 MW; A736998BDFAE34E2 CRC64;
SQ

```

Query Match 7.7%; Score 98.5; DB 1; Length 435;

Best Local Similarity 22.4%; Pred. No. 0.21; Indels 47; Gaps 8;

```

Matches 52; Conservative 42; Mismatches 91; Indels 47; Gaps 8;
QY 7 ERLNVLGGAAGKIAHIGVAKINE--IGIRRALSGVSGAATVYFAGSYSGMF 63
DB 209 EGVVLEAGTISRGTGNGFTYANKKESISLGI-----GCLVSDFOKGTETVYG- 257
QY 64 SLKRVNMLKLFKEPKLKIGWEKAIKRFLEVLTP--YRIEKLPIFYICATDLYSG 120
DB 258 -----LEBKSHPSVAPLIEGSEVVEVAHLIPGCGKALPOLFGDGMVVVGAQQL 309
QY 121 RALVYLSGSLIPALIGSCAIPGIF-----EPYKXNYLLVDGIVNNLPVEPFOESGIP 174
DB 310 NNAVREESNLAMTSGRIAAEALIVQYSRBPMSAKLL-----SLKTYLSESFV- 359
QY 175 TVCVVULPIBPEKDIKNILHLSFPLAV-RSNSEKREKFCOLVIVPELAE 225
DB 360 -----LKQMKKKYKDLPALHINSQNFILTYPOLVSKAMQNFVAVDGPVKYK 406

```

Search completed: June 19, 2003, 15:03:01

Job time : 23 secs

Thu Jun 19 17:30:47 2003

us-09-903-410-36.rpt

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 19, 2003, 14:50:55 ; Search time 44 Seconds

(without alignments)
1175,405 Million cell updates/sec

Title: US-09-903-410-36

Perfect score: 1273

Sequence: 1 LRKRFBENIVLSGGAAKNG.....RKAOIMERGVIKALVLS 251

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virius:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	934	73.4	259	16	067393
2	391.5	30.8	728	16	09HY6 aquifex seo
3	322.5	25.3	287	16	08HY6 pseudomonas
4	300.5	23.6	275	16	08HY6 thermomanaer
5	297.5	23.4	764	16	09K98 bacillus ha
6	294	23.1	319	16	09K98 vibrio chol
7	289	22.7	760	16	08HY6 ralteonias
8	287.5	22.6	304	2	09HY6 fuscobacteri
9	285.5	22.4	491	2	09HY6 erwinia chr
10	284.5	22.3	303	2	08HY6 bacteroides
11	284.5	22.3	321	16	09HY6 peptobacter
12	279	21.9	286	2	09HY6 rhizobium m
13	276	21.7	293	16	09HY6 vibrio mari
14	274.5	21.6	300	16	09HY6 neisseria m
15	273	21.4	314	16	08HY6 brucella me
16	273	21.4	315	16	08HY6 agrobacteri

17	268.5	21.1	297	16	09JXB2
18	268.5	21.1	395	16	09PH78
19	261.5	20.5	301	16	08EP44
20	259	20.3	289	16	09K3W7
21	258.5	20.3	314	16	08KDP1
22	257.5	20.2	301	16	08Z7F2
23	257	20.2	324	16	09BLQ4
24	253.5	19.9	148	16	005884
25	251.5	19.8	289	2	09CRD6
26	248	19.5	285	16	09CRD6
27	247.5	19.4	1065	16	08VJ44
28	243	19.1	345	16	09J185
29	240	18.9	311	16	08J196
30	238	18.7	303	16	08Y2M1
31	235.5	18.5	610	16	09A6C3
32	233.5	18.3	466	11	08R064
33	233.5	18.3	1240	11	08R3C5
34	232.5	18.3	630	4	08YAY5
35	232.5	18.3	702	4	096N75
36	231.5	18.2	311	16	08UHJ1
37	231	18.1	1071	16	069695
38	228	17.9	905	5	08SYN8
39	223.5	17.6	1327	4	060859
40	223.5	17.6	1327	11	09R114
41	208.5	16.4	1371	5	095PX1
42	199.5	15.7	296	16	09K945
43	176.5	13.9	290	12	091F63
44	172.5	13.6	382	4	09UG58
45	162	12.7	283	16	097GE3

ALIGNMENTS

RESULT 1	ID	067393	PRELIMINARY;	PRT;	259 AA.
AC	067393	01-AUG-1998 (T-REMBLrel. 07, Created)			
DT	01-AUG-1998 (T-REMBLrel. 07, Last sequence update)				
DT	01-JUN-2002 (T-REMBLrel. 21, Last annotation update)				
DB	Hypothetical protein A0_1386.				
GN	A0_1386.				
OS	Aquifex aeolicus.				
OC	Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.				
OX	NCBI_Taxid=63363;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=VFS;				
RX	MEDLINE=98196666; PubMed=9537320;				
RA	Decker G., Warren P.V., Gaasterland T., Young W.G., Jenox A.L.,				
RA	Graham D.E., Overbeek R., Sneed M.A., Keller M., Aulay M., Huber R.,				
RA	Feldman R.A., Short J.M., Olson G.J., Swanson R.V.,				
RT	"The complete genome of the hyperthermophilic bacterium Aquifex				
RT	aeolicus."				
RL	Nature 392:353-358(1998).				
DR	EMBL; AB000737; AAC07357.1; -				
DR	InterPro; IPR002641; Patatin.				
KW	Patam; Pf01734; Patatin; 1.				
KW	Hypothetical protein; Complete proteome.				
SQ	SEQUENCE 259 AA; 29099 MW; 91C2B986255B6501 CRC64;				
Query Match	Best Local Similarity	73.4%;	Score 934;	DB 16;	Length 259;
Matches	176;	Conservative 31;	Mismatches 40;	Indels 0;	Gaps 0;
QY	1 LRKRFBENIVLSGGAAKNGIHTVAKINELGIRPAISGVSAGATVVFASGYSB 60				
DB	5 LKRRFBENIVLSGGAAKNGIHTVAKINELGIRPAISGVSAGATVVFASGYSB 64				
QY	61 GMSLLKRVNWLGLTFKFKPLGLIGWEKATRFEEVLPYRREKLEIPYICATPLVYG 120				
DB	65 EMLKTLKRVNWLGLTFKFKPKKGLIGWEKATRFEEVLPYRREKLEIPYICATPLVYG 124				

DR InterPro: IPR002641; Patatin.
 DR InterPro: IPR001423; UPF0028.
 DR Pfam: PF01734; Patatin. 1.
 DR PROSITE: PS01237; UPF0028; 1.
 DR Hypothetical protein; Complete proteome.
 SQ SEQUENCE 275 AA; 30545 MW; 617C790E9AE7553 CRC64;

Query Match 23.6%; Score 300.5; DB 16; Length 275;
 Best Local Similarity 31.4%; Pred. No. 3.2e-19;
 Matches 71; Conservative 48; Mismatches 100; Indels 7; Gaps 3;

QY 8 EYNLVSGAAGKAIHIGVAKAINEIGIRVRLSGVSAAGIYVSVYASGSPKMF---S 64
 DB 8 KIGLALSGGARGRAHIGVAKVLEQEKIPIDYLAGSSKALVSAIYGAHGTETELIRAN 67
 QY 65 LIAKRVNMLKFKPKPKGLIGWEKAIPLBEVLPRRIEKLPIFYICATDLYSGALY 124
 DB 68 LFKRYKYL---DFVVPKRGFLAGHRYELIRYAKKKRYEELDPVRIVAADLTKGRVI 124
 QY 125 LSESGSLIPALISCAIPGIFEPVEYKNYLLVDGIVNNLPVPEPQSGIP--TVCVYLPI 183
 DB 125 LQEGVAVARASAIATIGIFVPKINIDRLIDGVIDRVPVGVVEMGADLTIVADLTIF 164
 QY 164 EBEKIKNIHLILRSFPLAVRSSEKKEFCULVPELEPFTL 229
 DB 185 REELIRSVYVILQITDMMSKELVRVQIDCTVMIRPNDRYSL 230

RESULT 5

Q9KUB9 PRELIMINARY; PRT; 764 AA.

AC O9KUB9; 09KUB9; 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein VCO603.
 GN VCO603.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heideberg U.F., Bisen U.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ertolaeva M.D., Vamathevan J., Bass S., Qin H., Diragol I., Sellers P.,
 RA McDonald L., Ueberback T., Fleischmann R.D., Niemann W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.D., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 RT cholerae.";
 RT Nature 406:477-483(2000).
 RU EMBL: A6004145; AAF93770.1; --
 DR TIGR; VCO603; --
 DR InterPro: IPR002106; AALRNA_1igaaseII.
 DR InterPro: IPR002641; Patatin.
 DR Pfam: PF01734; Patatin. 1.
 DR PROSITE: PS00339; AA TRNA LIGASE II 2; UNKNOWN_1.
 DR Hypothetical protein; Complete proteome.
 KW SEQUENCE 764 AA; 85616 MW; 0C3C8BCA100C840 CRC64;

Query Match 23.4%; Score 297.5; DB 16; Length 764;
 Best Local Similarity 27.6%; Pred. No. 2.2e-18;
 Matches 79; Conservative 61; Mismatches 105; Indels 41; Gaps 7;

QY 2 RLKREELNVLVSGAAGKAIHIGVAKAINEIGIRVRLSGVSAAGIYVSVYASGSPBG 61
 DB 28 QVAKRPIALVLAGGARGAAGHMGVLRALBEMHVPVDIITGSMGAVYGGVATQMSAE 87
 QY 62 MFSILKRVVNMVLEK-----FKPPLKGLIGWEKA 90

DB 88 IEALIVSDMNGYRDVDSQRAVRDKEYEDRYQITDGLHWEKVRAP--KGVYQGNM 146
 QY 91 IRLERV---LP-YRIEKLKIPFYICATDLYSGRALYISGSLIPALISCAIPGIFEP 146
 DB 147 LRLKRETNLPAFSPFQVLPYRAVATDIIHQEVLIDKGLVDMAMASVFGALP 206
 QY 147 VEYKNYLLVDGIVNNLPVPEPQSGIP--TVCVYL--LPIBEKIKNIHLILRSFPLAV 204
 DB 207 YEIDGMLVDGVTNNMVEVARAMGADIIVADISTDYSCDEFTNLTVAQDSNLY 266
 QY 205 RSNSEKKEFC---DVIVPELEPFTLVDVAKADQIMERGYAL 247
 DB 267 RSTERQSDHLTSRDLLRFPVGKEMTEPDKMPARAPMYQDAMD 312

RESULT 6

Q8Y066 PRELIMINARY; PRT; 319 AA.

AC Q8Y066; 08Y066; 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Probable lipoprotein transmembrane.
 GN RSC1178 OR RS04543.
 OS Ralstonia solanacearum (Pseudomonas solanacearum). 247
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 OC Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Arelat M., Billault A., Brothier P., Camus J.C., Catolico L.,
 RA Chander M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
 RT Nature 415:497-502(2002).
 RU EMBL: A646063; CAD14880.1; --
 DR InterPro: IPR002641; Patatin.
 DR Pfam: PF01734; Patatin. 1.
 KW Complete proteome.
 SQ SEQUENCE 319 AA; 32967 MW; 6924E744D4F53342 CRC64;

Query Match 23.1%; Score 294; DB 16; Length 319;
 Best Local Similarity 32.4%; Pred. No. 1.5e-18;
 Matches 84; Conservative 51; Mismatches 100; Indels 24; Gaps 8;

QY 9 INVLVSGAAGKAIHIGVAKAINEIGIRVRLSGVSAAGIYVSVYASGSP--EGMFSILK 67
 DB 53 IGLALSGGARGRAHIGVAKVLEQEKIPIDYLAGSSKALVSAIYGAHGTETELIRAN 112
 QY 68 R-----VNMVLEK--FKPPLKGLIGWEKAIPLBEVLPRRIEKLPIFYICATDLYSG 120
 DB 113 MDEKTTADMALPFGTGFSGMLG---BALRKVYNRVVKOKTIRBWKETPIGIVADLSSG 168
 QY 121 RALYLSGSLIPALISCAIPGIFEPVEYKNYLLVDGIVNNLPVPEPQSGIPVVCVYL 180
 DB 169 RPIIFRGNITGAVRASGIPGVFPVITSGHQVVDGLVAPVYVAKMGATFVIAVN 228
 QY 181 LPIEPE-KDKIKNIHLILRSFPLAVRSSEKKEFCULVPELEPFTLVDVAKADQIMERGYAL 234
 DB 229 ISAPSKQAVSGASMLDTTITMOSINKIELAQADVIVAFSLPFGVGSDFT---ARN 284
 QY 235 DQIM--ERGYKALEVLS 251
 DB 285 EALIGGEQALAMPLIRE 303

RESULT 7

Q8R6F6

QY 206 SNSKRRKPCDLYVPELEETPLDVKADQIMERGYTAKLEVL 249
 Db 264 KTDYSLPDSAGIIMTFKXNDVSLMDFORIDELKIGYDRTMSLM 307

RESULT 10
 ID 086196 PRELIMINARY; PRT; 303 AA.
 AC 086196;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE ExpL protein.
 GN ExpL.
 OS Pectobacterium carotovorum subsp. carotovorum.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Pectobacterium.
 CX NCBI_TaxID=555;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=SCC3193;
 RA Andersson R.A., Palva T.E., Pirhonen M.;
 RT "The response regulator ExpM is essential for the virulence of Erwinia
 RT carotovora subsp. carotovora and acts negatively on the sigma factor
 RT Rpos.";
 RL Mol. Plant Microbe Interact. 0:0-0(0).
 DR EMBL, AJ224437; CA11941.1;
 DR InterPro; IPR002641; Patatin.
 DR Pfam; PF01734; Patatin; 1.
 SQ SEQUENCE 303 AA; 33406 MW; FE972980ACAD1F9 CRC64;

Query Match 22.3%; Score 284.5; DB 2; Length 303;
 Best Local Similarity 28.3%; Pred. No. 1e-17; Indels 61; Gaps 7;
 Matches 82; Conservative 46; Mismatches 101;

QY 7 BEINIVLSGGAAGIAHIGVLKAINELGIRVALSGVAGAVSVFYASGYSPBEGMFSLL 66
 Db 4 KKTGIALSGGAAGIAHIGVLKAINELGIRVALSGVAGAVSVFYASGYSPBEGMFSLL 56

QY 67 KRVNMLKLPFKRPPLK-----GLIGWEKAIPLLEEVLPYRIEKLIPFYICADL 117
 Db 57 SMEQWVRGFGYWDVIRLMDLSWORGSLRSGDRVFNVSVKLTATTCIECALKYGVVTNL 116

QY 118 YSGRAIYLSGGSLIPALLGSCAIPGIFEEVEYKIVLVGGIVNNLPVEPFGESGIPYVC 177
 Db 117 STGRRLMTLEGDLHQWRASCSMPGLSPFRFNDYLVDAVAVNPVPSILAPAMGADIVI 176

QY 178 V-----DVLPREP-----EKDI-----KNILHLRL----- 198
 Db 177 AVDLQHDASINMODLSIKETYSVIDIMEHVSDQWRSIRRELLRGRQASBPPTAME 236

QY 199 --SFLVAVNSSEKKEPFC---DLVIVPELEETPLDVKADQIMERGY 242
 Db 237 IMSTSIQIENRLKMTKRMAGDPDVLLOPYCPQIATLDHRAQEALEAGY 286

RESULT 11
 ID 092086 PRELIMINARY; PRT; 321 AA.
 AC 092086;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical protein K01418.
 GN R01418 OR SMC01003.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Rhizobiaceae; Sinorhizobium.
 CX NCBI_TaxID=382;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;

RX MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Boisard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaune V., Masny D.,
 RA Fohl T., Portetelle D., Puehler A., Purnelle B., Ransperger U.,
 RA Renard C., Thebaud P., Vandendol M., Weidner S., Gallbert F.,
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 DR EMBL, AF591787; CAC45997.1;
 DR InterPro; IPR002641; Patatin.
 DR Pfam; PF01734; Patatin; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 321 AA; 34380 MW; D9AD7590B84D709F CRC64;

Query Match 22.3%; Score 284.5; DB 16; Length 321;
 Best Local Similarity 30.4%; Pred. No. 1e-17;
 Matches 86; Conservative 47; Mismatches 99; Indels 51; Gaps 6;

QY 9 INLVLSGGAAGIAHIGVLKAINELGIRVALSGVAGAVSVFYASGYSP---GMFS 64
 Db 40 IALALGGGAAGIAHIGVLKAINELGIRVALSGVAGAVSVFYASGYSP---GMFS 99

QY 65 LKRVNMLKLPFKRPPLKLGIGWEKAIPLLEEVLPYRIEKLIPFYICATDLYSGRALY 124
 Db 100 TVRRI--AGILDPALGGGIFGGLRLTKMQEHLQMLSELDLDPFVAVATEVYSGHEW 157

QY 125 LSEGSLIPALLGSCAIPGIFEEVEYKIVLVGGIVNNLPVEPFGESGIPYVC-----VD 179
 Db 158 IEGSLITAIRASYALPGIFEPNAGRTLIDGALVNPV-----SVGRABEQH 208

QY 180 VLPREPEDI-----KNILHLRSPFLAVNSSEK 210
 Db 209 VVANINLVYDVYGRSAVVKSHAGMETPDAPATBANSARLKMYSVWQAFNIQDRISBA 268

QY 211 R--KEPDDYIVPELEETPLDVKADQIMERGYTAKLEVL 251
 Db 269 RLADPDPDLIHLPLNDIGLSEPHRIGAEIERGIOEAKTYLSE 311

RESULT 12
 ID 09RA24 PRELIMINARY; PRT; 286 AA.
 AC 09RA24;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Genes, similar to eicosapentaenoic acid synthetase gene cluster.
 OS Vibrio marinus (Moritella marina).
 OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;
 CC Moritella.
 CX NCBI_TaxID=90736;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=MP-1;
 RA Tanaka M., Ueno A., Kawasaki K., Yumoto I., Ohgiya S., Hoshino T.,
 RA Ishizaki K., Okuyama H., Morita N.;
 RT "Isolation of clustered genes that are notably homologous to the
 RT eicosapentaenoic acid biosynthesis gene cluster from the
 RT docosahexaenoic acid-producing bacterium Vibrio marinus strain MP-1.";
 RL Biotechnol. Lett. 21:939-945(1999).
 DR EMBL, AB025342; BAA69379.1;
 DR InterPro; IPR002641; Patatin.
 DR Pfam; PF01734; Patatin; 1.
 SQ SEQUENCE 286 AA; 31587 MW; AAF729C1D2CE6C9 CRC64;

Query Match 21.9%; Score 279; DB 2; Length 286;
 Best Local Similarity 32.9%; Pred. No. 3e-17; Indels 12; Gaps 5;
 Matches 80; Conservative 42; Mismatches 109;

QY 8 EINIIVLSGGAAGIAHIGVLKAINELGIRVALSGVAGAVSVFYASGYSPBEGMFS 64
 Db 7 KIGIVLSGGAAGIAHIGVLKAITLEODIRPNVIACTSGSNVAGALYSGLEIDLOFFI 66

RA Selkov B., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
 RA Haselkorn R., Kyriides N. Overbeek R.;
 RT "The genome sequence of the facultative intracellular pathogen
 RT Brucella melitensis." U.S.A. 99:443-448(2002).
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 DR EMBL; AE009531; AAL52089.1; -
 DR InterPro; IPR002641; Patatin.
 DR Pfam; PF01734; Patatin; 1.
 KM Hydrolase; Complete proteome.
 SQ SEQUENCE 314 AA; 34266 MW; E18ADC1225AFDB CRC64;

Query Match 21.4%; Score 273; DB 16; Length 314;
 Best Local Similarity 27.8%; Pred No. 1.2e-16;
 Matches 80; Conservative 52; Mismatches 98; Indels 58; Gaps 7;
 QY 7 BEINLVISGSAKGIHIGVLKAINELGIRVALSGVSAGAIVSIFYAGYSPE----- 60
 DB 33 QKIALALGGGAAGMWHIGVLRALDAGIEIEMIAGTISIGALVGGCYLAGKNELEEFAR 92
 QY 61 -----GWFSLAKRVNMLKLFKFPPLKGLIGMEKAIKFLSEVLPIYRIEKLBIPTYICAT 115
 DB 93 SLTRRRMFNLD-----ITPRG--SGLPQMKLDGRLREHLDGIRIEDLPKPFVAVCT 143
 QY 116 DLYSGRALYISBSGLIPALGSCAIPGIFEEVEYKNTLYVDGIVNNLPY----- 166
 DB 144 ELKGTGHEIWLSTGPLEANRASYALPGVFEPVWGDRLVDGALVNPVPSVCRAVEORL 203
 QY 167 -----PQESGIFPTVCVDVLP1BPEKDIK-NILHILLRSFFLAVR 205
 DB 204 VLAAHLHYDYGRAAVIKAKMRORSVPOA-----MHGSEKTRLGITGVMMEAFFNIID 257
 QY 206 SNSEKPK--EPCDLVIVPELEEFPTLDVRAKADQIMERGYIKALEVLSB 251
 DB 258 RISRAIRAGDPDPVSLMPTVGQIGLADFHRAAEALDAGYETETVRIED 305

Search completed: June 19, 2003, 15:04:02
 Job time : 45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 12:47:00 ; Search time 173 Seconds
(without alignments)
9841.102 Million cell updates/sec

Title: US-09-903-410-26

Perfect score: 756
Sequence: 1 ttcagattcgaggaatttga.....tagagcttcttcgatatg 756

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq, 101002:*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	752.8	99.6	756	18	AA179324
2	396.2	52.4	750	18	AA179327
3	53.6	7.1	903	20	AA121214
4	52	6.9	903	20	AA121213
5	52	6.9	92934	21	AA181473
6	52	6.9	172325	21	AA121613
7	52	6.9	837096	21	AA181489
8	49.8	6.6	1877	23	AA179447
9	47.4	6.3	903	20	AA121215

10	46.2	6.1	40138	20	AA181946	V. marinus PK8-lik
11	46.2	6.1	40138	21	AA171520	V. marinus PK8-lik
12	46.2	6.1	41587	22	AA164984	Moritella maritima
13	45.8	5.1	467	24	ABK79457	Bacillus clausii g
14	42.6	5.6	447	20	AA121212	Neisseria meningit
15	42.6	5.6	447	21	AA181350	N. meningitidis Me
16	39.4	5.2	269223	22	AA128554	Genomic fragment #
17	39.4	5.2	12019	20	AA181867	Alcaligenes sp. Po
18	36.6	4.8	1497	22	AA161015	P. putida KT240-a
19	35.8	4.7	876	24	ABK75059	Bacillus lichenifo
20	35.4	4.7	5591	23	AB104293	Drosophila melanog
21	35.4	4.7	13437	23	AB104292	Drosophila melanog
22	35	4.6	1484	21	AA147412	Arabidopsis thalia
23	35	4.6	1487	21	AA136225	Arabidopsis thalia
24	34.4	4.6	11628	20	AA120520	Polynucleotide seq
25	34	4.5	1299	20	AA120681	Polynucleotide seq
26	33.2	4.4	5520	19	AA103311	Thermococcus 9N-2
27	33	4.4	472	22	AA184381	E. coli growth and
28	33	4.4	10944	22	AA146250	DNA encoding novel
29	32.8	4.3	3793	24	ABK92234	Prostate cancer-as
30	32.8	4.3	5948	23	AB114817	Drosophila melanog
31	32.8	4.3	11046	23	AB114816	Drosophila melanog
32	32.8	4.3	34980	22	AA141223	Pyrococcus abyssi
33	32.6	4.3	34980	22	AA141224	Pyrococcus abyssi
34	32.6	4.3	494	22	AA136495	Probe #5181 used t
35	32.6	4.3	494	22	ABK50206	Human genome-deriv
36	32.6	4.3	567	24	ABK53343	Human eosinophil-m
37	32.4	4.3	994	24	ABQ69148	Listeria monocytog
38	32.4	4.3	994	24	ABQ70659	Listeria monocytog
39	32.4	4.3	2944528	24	AB103041	Listeria monocytog
40	32.2	4.3	4403765	22	AA199683	Mycobacterium tube
41	32.2	4.3	4411529	22	AA199682	Mycobacterium tube
42	32	4.2	811	24	AA199682	Human prostate exp
43	32	4.2	1476	21	AA147228	Arabidopsis thalia
44	32	4.2	1478	21	AA142167	Arabidopsis thalia
45	32	4.2	17569	24	ABK39756	cDNA encoding clon

ALIGNMENTS

RESULT 1
AA179324 standard, DNA, 756 BP.

AA179324;

16-FEB-1998 (first entry)

DNA encoding Aquifex pyrophilus esterase 28LC.

Esterase; thermostable enzyme; ester; chiral compound; cheese;
pulp; paper; lignin removal; sugar; lignocellulose;
disease resistance; feedstuff; ss.

Aquifex pyrophilus strain KO1 5a.

Key Location/Qualifiers

FT CDS 1..756 /*tag= a

FT /transl_except= (pos: 742..746, aa:Val)

XX WO9730160-A1.

XX PD 21-AUG-1997.

XX PF 11-FEB-1997; 97WO-US02039.

XX PR 16-FEB-1996; 96US-0602359.

XX PA (RECO-) RECOMBINANT BIOTRANSFORMATION INC.

XX PI Callen W, Kosmicka A, Link S, Maffia AM, Murphy D;

Reid J, Robertson DE, Swanson RV, Warren PV;
 WPI; 1997-425035/39.
 P-PSDB; AAW23071.

Nucleic acid encoding heat stable esterase from thermophilic bacteria - which is active in organic solvents, useful in cheese or paper manufacture, and to study plant resistance to disease

Claim 1; Page 44-45; 113pp; English.

This DNA sequence codes for thermostable esterase 28LC (AAW23071) of Aquifex pyrophilus, a Gram-negative, chemolithoautotrophic knall gas marine bacterium which grows optimally at 85 deg C and pH 6.8. It can be amplified from a pBluescript vector by PCR (see AAT79306-07). Claimed, newly identified polynucleotides (AAT79321-30) encoding esterases (AAW23069-77, AAW23088) were recovered from genomic gene libraries. They can be used for recombinant production of the enzymes in host cells, and as probes to identify related sequences. The esterases are stable at high temperature and in organic solvents, making them superior for use in pharmaceutical, agricultural and other chemical industries. A method is claimed for transferring an amino group from an amino acid to an alpha-keto acid using a claimed esterase.

The enzymes may also be useful as ripening starters in cheese making, in lignin removal in paper and pulp manufacture, in carbohydrate derivative synthesis, in fermentable sugar production from lignocellulosic waste, in the study of plant wall structure, plant resistance to disease and organic matter decomposition and to select plants bred for production of highly degradable animal feeds.

Sequence 756 BP; 202 A; 153 C; 201 G; 200 T; 0 other;

Query Match 99.6%; Score 752.8; DB 18; Length 756;
 Best Local Similarity 99.7%; Pred. No. 3e-239;
 Matches 754; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 TTGAGATTGAGAAATTGAAGAGATAAACCCTCGTTCTTCGCGAGGAGCTGCAAGGGC 60
 1 TTGAGATTGAGAAATTGAAGAGATAAACCCTCGTTCTTCGCGAGGAGCTGCAAGGGC 60
 61 ATAGCCCATATAGTGTGTTTGAAGAGCTATAACAGAGCTCGGTATTAAGGTCGAGGCTTAA 120
 61 ATAGCCCATATAGTGTGTTTGAAGAGCTATAACAGAGCTCGGTATTAAGGTCGAGGCTTAA 120
 121 AGCGGGTGAGCGCCGGGGCAATCGTTTCGGTCTTTTATGCTCAGGCTACTCCCTGAA 180
 121 AGCGGGTGAGCGCCGGGGCAATCGTTTCGGTCTTTTATGCTCAGGCTACTCCCTGAA 180
 181 GGGATGTTTTCAGCTTCTGAAGAGGGTAACTCGTGTGTTTATGCTCAGGCTACTCCCTGAA 240
 181 GGGATGTTTTCAGCTTCTGAAGAGGGTAACTCGTGTGTTTATGCTCAGGCTACTCCCTGAA 240
 241 CTGAAGGATTTGATAGGTCGGGAGAGGCTATAAGATCTCTTGAGGAAAGTTCTCCCTTAC 300
 241 CTGAAGGATTTGATAGGTCGGGAGAGGCTATAAGATCTCTTGAGGAAAGTTCTCCCTTAC 300
 301 AGAGAAATAGAAAACCTTGAGATACCGACGATATATATGCGGACGATTTATCTCGGGA 360
 301 AGAGAAATAGAAAACCTTGAGATACCGACGATATATATGCGGACGATTTATCTCGGGA 360
 361 AGGGCTCTATACCTCTCGGAAGGAGTGTAAATCCCGCACTTCTCGGACAGCTGTGCAATT 420
 361 AGGGCTCTATACCTCTCGGAAGGAGTGTAAATCCCGCACTTCTCGGACAGCTGTGCAATT 420
 421 CCCGGCATATTGAAACCCGTTGAGTATAAGAAATTAATCTGCTGTTGACGAGGTATAGTT 480
 421 CCCGGCATATTGAAACCCGTTGAGTATAAGAAATTAATCTGCTGTTGACGAGGTATAGTT 480
 481 AACAACTTCCCTGAGCCCTTTACAGGAAGCGGTATTTCCACCGCTTTGCTGCTATGTC 540
 481 AACAACTTCCCTGAGCCCTTTACAGGAAGCGGTATTTCCACCGCTTTGCTGCTATGTC 540

541 CTTCCATAGAGCCGGAAGAGATATATAAGAACATTTCTTACATCCTTTTGAGGAGCTTC 600
 541 CTTCCATAGAGCCGGAAGAGATATATAAGAACATTTCTTACATCCTTTTGAGGAGCTTC 600
 601 TTTCTTGGCTCCGCTCAAACTCCGAAAGAGAGAGTGTGTTGACCTCGTTATAGTT 660
 601 TTTCTTGGCTCCGCTCAAACTCCGAAAGAGAGAGTGTGTTGACCTCGTTATAGTT 660
 661 CTTGAGCTTGGAGAGTTTACACCCCTTCATGTTTAAAAAGCGGACCAATTAATGAGAGG 720
 661 CTTGAGCTTGGAGAGTTTACACCCCTTCATGTTTAAAAAGCGGACCAATTAATGAGAGG 720
 721 GGATACATTAAGGCTTGAAGAGTACTTCTGAATAG 756
 721 GGATACATTAAGGCTTGAAGAGTACTTCTGAATAG 756

RESULT 2
 AAT79327
 ID AAT79327 standard; DNA; 750 BP.
 XX
 AC AAT79327;
 XX
 DT 16-FEB-1998 (first entry)
 XX
 DE DNA encoding Aquifex esterase VF5-34LC.
 XX
 KW Esterase; thermostable enzyme; ester; chiral compound; cheese;
 KW pulp; paper; lignin removal; sugar; lignocellulose;
 KW disease resistance; feedstuff; ss.
 XX
 OS Aquifex sp. strain VF5.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..750
 FT /tag= a
 FT /transl_except= (pos: 295..297, aa:Glu)
 XX
 PN WO9730160-A1.
 XX
 PD 21-AUG-1997.
 XX
 PF 11-FEB-1997; 97MO-US02039.
 XX
 PR 16-FEB-1996; 96US-0602359.
 XX
 PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
 XX
 PI Callen W, Kosmotka A, Link S, Maffia AM, Murphy D;
 PI Reid J, Robertson DE, Swanson RV, Warren PV;
 XX
 XX WPI; 1997-425035/39.
 DR P-PSDB; AAW23071.
 XX
 XX Nucleic acid encoding heat stable esterase from thermophilic
 XX bacteria - which is active in organic solvents, useful in cheese or
 XX paper manufacture, and to study plant resistance to disease
 XX
 XX Claim 1; Page 48-49; 113pp; English.
 XX
 XX This DNA sequence codes for thermostable esterase VF5-23LC
 XX (AAW23074) of Aquifex VF5, a marine strictly chemolithoautotrophic
 XX knall gas bacterium that grows optimally at 85 deg C and pH 6.8.
 XX It can be amplified from a pBluescript vector by PCR (see AAT79312-
 XX 13). Claimed, newly identified polynucleotides (AAT79321-30) encoding
 XX esterases (AAW23069-77, AAW23088) were recovered from genomic libraries.
 XX They can be used for recombinant production of the enzymes in host
 XX cells, and as probes to identify related sequences. The esterases
 XX are stable at high temperature and in organic solvents, making them
 XX superior for use in production of pure chiral compounds used in
 XX pharmaceutical, agricultural and other chemical industries. A
 XX method is claimed for transferring an amino group from an amino
 XX acid to an alpha-keto acid using a claimed esterase. The enzymes

QY 439 GTTGAAGTATAAGAAATTACTCTCTGTTGACCGAGGTATAGTTAAACACCTTCCTCCGT 494
 |||||
 Db 550 GTTATCATCGGCAGGCATACATATGTTGACCGCGTCTGTGCGACGCCGTCGCGCGT 605

RESULT 4

AAZ12213
 ID AAZ12213 standard; DNA; 903 BP.

AC AAZ12213;
 XX

DT 08-OCT-1999 (first entry)

DE Neisseria meningitidis complete ORF137 sequence.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea; ss.

XX Neisseria meningitidis.
 OS

PN WO9924578-A2.

PD 20-MAY-1999.

XX 09-OCT-1998; 98WO-IB01665.

PR 01-SEP-1998; 98GB-0019016.

PR 06-NOV-1997; 97GB-0023516.

PR 14-NOV-1997; 97GB-0024190.

PR 18-NOV-1997; 97GB-0024386.

PR 27-NOV-1997; 97GB-0025158.

PR 10-DEC-1997; 97GB-0026147.

PR 14-JAN-1998; 98GB-0000759.

XX (CHIR-) CHIRON SPA.

XX Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;

PI WPI; 1999-327407/27.

DR P-PSDB; AAY38778.

XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
 PT diagnosis, treatment and prevention of infection

XX Claim 9; Page 323; 524pp; English.

XX Nucleotide sequences AAZ11972-212358 represent open reading frames
 CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode
 CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their
 CC fragments, their nucleic acids and antibodies are used for diagnosis,
 CC prevention (as vaccines) or treatment of Neisseria infections,
 CC such as meningitis, septicaemia and gonorrhea. Both organisms
 CC are closely related. Fragments of the nucleic acids are useful
 CC as hybridisation probes and antisense reagents.

XX Sequence 903 BP; 204 A; 230 C; 255 G; 214 T; 0 other;

SQ Query Match 6.9%; Score 52; DB 20; Length 903;

Best Local Similarity 44.3%; Pred. No. 1.8e-06;

Matches 211; Conservative 0; Mismatches 265; Indels 0; Gaps 0;

QY 19 GAAGAGATAAACTCGTTCCTTCGCGAGGAGCTGCAAGGCATAGCCACATAGTGT 78

Db 130 GCATGTGCGTTTGGCATCTCGTGGCGGCATCTAAGGATTTGCCATGTAGTATT 189

QY 79 TTGAAGCTATAAACGAGCTCGGTATAAGGCTGAGGGCTTTAAGCGGGGTGAGCCGGG 138

Db 190 ATTAAGCTTTTGAAGAAAAACGGTATTCTCTGTGAAGGTGTTACCGGCACATCGGCAGGT 249

QY 139 GCATCGTTCGTTCTTTATGCTCAGGCTACTCCCTGAAGGGATGTTACGCTTCTG 198

Db 250 TCGATTGCGGCACCTTTTTCATCGGGTATGTCGCCGACCGCCTCGAATTTGNAGCC 309

QY 199 AAGAGGTAAACTGCTGAGCTGTTTAAAGTTCAAGCCACCTCTCAAGGGATTGATAGG 258
 |||||
 Db 310 GAAATTTTGGCAAAACCGATTGTGTCGATTTAAACCTTGTCCACAGTGGTTTATCAA 369
 |||||
 QY 259 TGGGGAAGGCTATAAGATTCTCTCAGGAAGTTCTCCCTTACAGGAGAAATAGAAAACCTT 318
 |||||
 Db 370 GCGGAAGCTCAGAAATTACATCAACCGAAAAGTCGGCGGCGAGCATTCAGCAGTTT 429
 |||||
 QY 319 GAGATACCGACGATATATATGCGCGAGGATTTATTAATCTGGGAGGGCTCTATACCTCTCG 378
 |||||
 Db 430 CCCATCAAAATTTGCGCGCGTTCCTACTGATTTTGAACCGCGCAAGCGCGTTCGCTTTCAAT 489
 |||||
 QY 379 GAAGGGGTTTAATCCCGCACTCTCGGCAGCTGTGCAATTCGCGCATATTTGAACCC 438
 |||||
 Db 490 CAGGGGATCGCGGCGAGCTGTGCGGCTTCGCGGCATTCCTCAATGTGTTCCACCC 549
 |||||
 QY 439 GTTGAATATAAGAAATTTACTTGTCTGTGACGGAGGTATAGTTAAACACCTTCCTCGT 494
 |||||
 Db 550 GTTATCATCGGCAGGCATACATATGTTGACGCGGTCTGTGCGACGCCGTCGCGCGT 605

RESULT 5

AAA81473/c

ID AAA81473 standard; DNA; 92934 BP.

XX AAA81473;

AC AAA81473;

XX 04-DEC-2000 (first entry)

XX N. meningitidis partial DNA sequence gnm_21 SEQ ID NO:21.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW Meningococcus B; MenB; ds.

XX Neisseria meningitidis.

XX WO200022430-A2.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-US23573.

XX 09-OCT-1998; 98US-0103794.

XX 30-APR-1999; 99US-0132068.

XX (CHIR) CHIRON CORP.

XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 XX Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 XX Rappuoli R, Pizza M;

XX WPI; 2000-318079/27.

XX Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisserial infections, for example, N.gonorrhoea -
 XX Claim 7; Page 471-498; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to

CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against *Meningococcus B*: against all serotypes;
 CC and/or against all pathogenic *Neisseriae*. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.

SO Sequence 92934 BP; 24051 A; 24249 C; 21902 G; 22730 T; 2 other;

Query Match 6.9%; Score 52; DB 21; Length 92934;
 Best Local Similarity 44.3%; Pred. No. 2.2e-05;
 Matches 211; Conservative 0; Mismatches 265; Indels 0; Gaps 0;

QY 19 GAAGAGATTAACCTGCTTCTTGGGAGAGCTGCAAGGCGATAGCCCATAGGTGTT 78
 DB 50133 GCGATGTCGCTTGGCACTCGGCGCGCATTAAGGATTTGCCATGTAGGTATT 50074
 QY 79 TTGAAGCTATTAACGAGCTCGGTATAGGGGTAGGGCTTTAAGCGGGTGAGCGCGG 138
 DB 50073 ATTAAGGTTTGAAGAAAGGATATCTGTGAAGGTGTACCGGCACATCGCGAGGT 50014
 QY 139 GCATCGTTCGCTCTTTATGCTCAGGCTACCTCCCTGAAGAGATTTACCTCTG 198
 DB 50013 TCGATTTTCGCGACCTTTTGTGATCGGGTATGTGCGCCGACCGCTCGAATTGAAAGCC 49954
 QY 199 AAGAGGTTAACTGCTGAGAGCTGTTAAGTTCAGCCACTCGAAGGATTTGATAGG 258
 DB 49953 GAAATTTTAGGCAAAACCGATTTGTGATTAACCTTTGACACAGTGTATTATCAA 49894
 QY 259 TGGAGAGAGCTATTAAGATTCCTTGAAGAAATTCCTCTTACAGAGAAATGAAAACTT 318
 DB 49893 GGGCAAAAGCTGCAAAATTAATCATCACGAAAGTGGCGGAGGAGATTCAGAGTTT 49834
 QY 319 GAGATACCGAGCTATTAATATGCGGAGGATTTATCTCGGAAAGGGCTCTATACCTCTG 378
 DB 49833 CCCATCAAAATTTGCCCGCTGTGCTACTGATTTTGAACCGCGAAGCGCGTTCAT 49774
 QY 379 GAAGGAGTTTAATCCCGCACTTCGCGAGCTGTGCAATTCGCCGATATTGAAGCC 438
 DB 49773 CAGGGAAATGCGGCGAGCTGTGCGGCTTCGCGCGCATTCGCAATGTGTCAACCC 49714
 QY 439 GTTGAATTAAGATTAATCTGCTGCTGTAGCGAGGTATTAATTAACAATTCCTCCGT 494
 DB 49713 GTTATCATCGGCAAGCATATATGTTGACGGCGGTCTGTGCGAGCGCGTCCCGT 49658

RESULT 6
 AAF21613/c
 ID AAF21613 standard; DNA; 172325 BP.

XX AAF21613;
 DT 13-MAR-2001 (first entry)
 DE *Neisseria meningitidis B* nucleotide sequence SEQ ID NO:114.
 XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; immunogenic; vaccine;
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
 KW ds.
 XX *Neisseria meningitidis*.
 OS
 XX MO20006791-A1.
 PN
 XX 09-NOV-2000.
 PF 08-MAR-2000; 2000WO-US05928.

XX 30-APR-1999; 99US-0132068.
 PR 08-OCT-1999; 99WO-US23573.
 PR 28-FEB-2000; 2000GB-0004695.
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 PI Pizsa M, Hickey E, Peterson J, Tettelin H, Venter JC, Maignani V,
 PI Galocci C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R,
 PI Frazer CM, Grandi G;
 DR WPI; 2000-647603/62.

PT *Neisseria meningitidis B* full length genome sequence and open reading
 PT frames are used to detect, treat and prevent *Neisserial* infections -
 PS Claim 7; Appendix A; 692pp; English.

CC The present invention describes the full length genome of
 CC *Neisseria meningitidis B* (NMB). The sequences in AAF21544 and AAF21607
 CC to AAF21613 represent fragments of the NMB genomic sequence, as the
 CC sequence was too long to go in a record on its own it was split into 8
 CC sequences which overlap each other at the beginning and end of each
 CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
 CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
 CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
 CC *Neisseria* proteins given in AAB58550 to AAB58593, and AAF21589 to
 CC AAF21606 represent PCR primers which are used in the exemplification of
 CC the present invention. The NMB genome and fragments from it have
 CC antibacterial activity, and can be used in vaccines and gene therapy.
 CC *Neisseria* nucleic acids, proteins and/or antibodies which binds to the
 CC proteins can be used in compositions for treating or preventing infection
 CC due to *Neisserial* bacteria or as a diagnostic reagent for detecting the
 CC presence of *Neisserial* bacteria or of antibodies raised to *Neisserial*
 CC bacteria. Computers, computer memory, computer storage medium or computer
 CC databases can be used in a search to identify open reading frames (ORFs)
 CC or coding sequences within the NMB genome. The DNA sequences provide
 CC further opportunities to find antigenic or immunogenic proteins which are
 CC more effective in vaccines than the outer membrane proteins currently
 CC used.

SO Sequence 172325 BP; 43072 A; 47583 C; 41465 G; 40205 T; 0 other;

Query Match 6.9%; Score 52; DB 21; Length 172325;
 Best Local Similarity 44.3%; Pred. No. 3e-05;
 Matches 211; Conservative 0; Mismatches 265; Indels 0; Gaps 0;

QY 19 GAAGAGATTAACCTGCTTCTTGGGAGAGCTGCAAGGCGCATAGCCCATAGGTGTT 78
 DB 153800 GCGATGTCGCTTGGCACTCGGCGCGCATTAAGGATTTGCCATGTAGGTATT 153741
 QY 79 TTGAAGCTATTAACGAGCTCGGTATAGGGGTAGGGCTTTAAGCGGGGTAGAGCGCGG 138
 DB 153740 ATTAAGGTTTGAAGAAAGGATATCTGTGAGAGGTGTACCGCGACATCGCGAGGT 153681
 QY 139 GCAATCGTTCGCTCTTTATAGCTCAGGCTACTCCCTGAAGGAGATTTGACCTCTG 198
 DB 153680 TCGATTTGCGGAGCTTTTGTGATCGGTATGTGCGCGACCGCTCGAATTGGAAGCC 153621
 QY 199 AAGAGGTTAACTGCTGAGAGCTGTTTAAGTTCAGCACTCTGAGAGGATTTAGAGG 258
 DB 153620 GAAATTTTAGGCAAAACCGATTTGTGATTAACCTTTGTCACAGAGTGTTTATCAA 153561
 QY 259 TGGAGAGAGCTATTAAGATTCCTTGAAGAAATTCCTCTTACAGAGAAATGAAAACTT 318
 DB 153560 GCGGAAAGCTGCAAAATTAATCAACGAAAGTGGCGGAGGACAGATTCAGAGTTT 153501
 QY 319 GAGATACCGAGCTATTAATATGCGGAGGATTTATCTCGGAAAGGGCTCTATACCTCTG 378
 DB 153500 CCCATCAAAATTTGCCCGCTGTGCTACTGATTTTGAACCGGAGAGGCGCTTTCAT 153441
 QY 379 GAAGGAGTTTAATCCCGCACTTCGCGAGCTGTGCAATTCGCGCATATTGAAGCC 438

Db	153440	CAGGGGAATGCCGGCAGGCTGCGCGCTTCGCCGCCATCCCAATGTTCACACC	153381
Qy	439	GTTCAGTATAGAATTAATCTGCTGCTGACGGAGGTATAGTTAAACACTTCCCGT	494
Db	153380	GTTATCATCGCAGGCATACATATGTTGACGGCGGTCTGTGCGAGCCCGTGC	153325
RESULT 7			
AAAB1489/c			
ID	AAAB1489	standard; DNA; 837096 BP.	
XX	AAAB1489;		
XX	04-DEC-2000	(first entry)	
DT	04-DEC-2000	(first entry)	
DE	N. meningitidis partial DNA sequence gnm_37	SEQ ID NO:37.	
KW	Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;		
KW	antigen; vaccine; diagnosis; infection; antibacterial; identification;		
KW	Meningococcus B; MenB; ds.		
OS	Neisseria meningitidis.		
XX	WO200022430-A2.		
XX	20-APR-2000.		
XX	08-OCT-1999;	99WO-US235573.	
XX	09-OCT-1998;	98US-0103794.	
PR	30-APR-1999;	99US-0132068.	
XX	(CHIR)	CHIRON CORP.	
XX	Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;		
PI	Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V;		
PI	Rappuoli R, Pizza M;		
XX	WPI; 2000-318079/27.		
DR	Isolated nucleotide sequences of Neisseria meningitidis which can be used in the diagnosis and treatment of N. meningitidis infection and other Neisserial infections, for example, N.gonorrhoea -		
XX	Claim 7; Page 629-865; 1760pp; English.		
XX	The present invention describes methods of obtaining immunogenic proteins from Neisseria genomic sequences. AAA81453 to AAA82414 represent specifically claimed Neisseria meningitidis genomic DNA sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent Neisseria DNA sequences and their corresponding proteins; AAA81254 to AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the isolation of Neisseria meningitidis DNA sequences; and AAA81322 to AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF sequences, which are all used in the exemplification of the present invention. The nucleic acid sequences, protein sequences, and antibodies against them, can be used in the manufacture of a composition. The composition can be used as a medicament (or in the manufacture of a medicament) for treating, preventing or diagnosing infection due to Neisserial bacteria. For example, some of the identified proteins could be components of vaccines against Meningococcus B; against all serotypes; and/or against all pathogenic Neisseriae. Identification of biological probes, from the bacterium will also facilitate production of biological probes, particularly organism-specific probes. Attempts to make efficacious Meningococcus B vaccines have failed mainly due to antigen tolerance. Multivalent vaccines have also been tried but none have successfully overcome antigenic variability. The provision of further, complete sequences may provide an opportunity to identify secreted or surface exposed proteins that may be presumed targets for the immune system and which are not antigenically variable or at least more conserved than other more variable regions.		

Db	153440	CAGGGGAATGCCGGGAGGCTGTGCGGCTTCGGCGGCATTCCTCAATGTGTTCCAAACC	153381
Qy	439	GTTGAGTATAAGAAATTAATCTGCTGCTGACGGAGGTATAGTTAAACAACTTCCTCGT	494
Db	153380	GTTATCATCGCAGGCATACATATGTTGACGGCGGTCTGTGCGCAGCCCGTGCCTCGT	153325

RESULT 7

AAAB1489/c

ID AAAB1489 standard; DNA; 837096 BP.

XX AAAB1489;

XX 04-DEC-2000 (first entry)

XX N. meningitidis partial DNA sequence gnm_37 SEQ ID NO:37.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;

KW antigen; vaccine; diagnosis; infection; antibacterial; identification;

KW Meningococcus B; MenB; ds.

XX Neisseria meningitidis.

XX WO200022430-A2.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-US23573.

XX 09-OCT-1998; 98US-0103794.

PR 30-APR-1999; 99US-0132068.

XX (CHTR) CHIRON CORP.

XX Frazer CM, Hickey B, Peterson J, Tettelin H, Venter JC;

PI Maslignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;

PI Rappuoli R, Pizza M;

XX WPI: 2000-318079/27.

DR

XX Isolated nucleotide sequences of Neisseria meningitidis which can be

PT used in the diagnosis and treatment of N. meningitidis infection and

PT other Neisserial infections, for example, N.gonorrhoea -

XX Claim 7; Page 629-865; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic

CC proteins from Neisseria genomic sequences. AAAB1453 to AAAB2414

CC represent specifically claimed Neisseria meningitidis genomic DNA

CC sequences; AAAB1260 to AAAB1303 and AAAB25620 to AAAB25663 represent

CC Neisseria DNA sequences and their corresponding proteins; AAAB1254 to

CC AAAB1259 and AAAB1304 to AAAB1321 represent PCR primers used in the

CC isolation of Neisseria meningitidis DNA sequences; and AAAB1322 to

CC AAAB1452 represent Neisseria meningitidis MenB polynucleotide ORF

CC sequences, which are all used in the exemplification of the present

CC invention. The nucleic acid sequences, protein sequences, and antibodies

CC against them, can be used in the manufacture of a composition. The

CC composition can be used as a medicament (or in the manufacture of a

CC medicament) for treating, preventing or diagnosing infection due to

CC Neisserial bacteria. For example, some of the identified proteins could

CC be components of vaccines against Meningococcus B; against all serotypes;

CC and/or against all pathogenic Neisseriae. Identification of sequences

CC from the bacterium will also facilitate production of biological probes,

CC particularly organism-specific probes. Attempts to make efficacious

CC Meningococcus B vaccines have failed mainly due to antigen tolerance.

CC Multivalent vaccines have also been tried but none have successfully

CC overcome antigenic variability. The provision of further, complete

CC sequences may provide an opportunity to identify secreted or surface

CC exposed proteins that may be presumed targets for the immune system and

CC which are not antigenically variable or at least more conserved than

CC other more variable regions.

XX

PT biodiversity -
 XX
 PS Claim 1, SEQ ID No 13751, 103bp; English.
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations in
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1877 BP; 426 A; 464 C; 516 G; 470 T; 1 other;
 Query Match 6.6%; Score 49.8; DB 23; Length 1877;
 Best Local Similarity 52.7%; Pred. No. 1.4e-05;
 Matches 108; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
 QY 340 GCGACGATTAATCTCGGAAAGGGCTCTACCTCGGAAAGGAGTTTATCCCGCA 399
 DB 54 GCCACCAATTAATAGTACGGAACGTAATATGTTTCTGAAGCAATCTTCATTTGCT 113
 QY 400 CTCTCGGAGAGCTGTGCAATTCGCGCATATTGAAACCGCTTGAGTAAAGAAATTA 459
 DB 114 ATTGCGCATCATCATGATTTCCAGACTCATGCACTGTGCACTAAACGGCTACTGG 173
 QY 460 CTGCTTACGAGAGTATAGTAAACACTTCCGTTAGCCCTTTACGAAAGCGTAT 519
 DB 174 CTGCTTATAGAGAGAGTGTAAACCAATTCCTTACCGCTGATTTGGGGCT 233
 QY 520 CCCACCGCTTTCGTTGATGTCCTTC 544
 DB 234 GATATTGTGATAGCGGTTGACCTGC 258
 RESULT 9
 AA212215
 ID AA212215 standard; DNA; 903 BP.
 XX
 XX AA212215;
 AC
 XX 08-OCT-1999 (first entry)
 DT
 XX
 DE Neisseria gonorrhoeae complete ORF137 sequence.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW treatment; Neisseria infection; meningitis; septicemia; gonorrhea; ss.
 XX
 OS Neisseria gonorrhoeae.
 XX
 PN MO9924578-A2.
 XX
 XX 20-MAY-1999.
 PD
 XX
 PF 09-OCT-1998; 98WO-IB01665.
 XX
 PR 01-SEP-1998; 98GB-0019016.
 PR 06-NOV-1997; 97GB-0023516.
 PR 14-NOV-1997; 97GB-0024190.

PR 18-NOV-1997; 97GB-0024386.
 PR 27-NOV-1997; 97GB-0025158.
 PR 10-DEC-1997; 97GB-0026147.
 PR 14-JAN-1998; 98GB-0000759.
 XX
 PA (CHIR-) CHIRON SPA.
 XX
 PI Grandi G, Maegnan V, Pizza M, Rappoli R, Scarlato V;
 DR MPI, 1999-327407/27.
 DR P-PSDB; AAV38780.
 XX
 PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
 XX diagnosis, treatment and prevention of infection
 PS Claim 9; Page 324-325; 524pp; English.
 CC Nucleotide sequences AA211972-213358 represent open reading frames
 CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode
 CC antigenic proteins (see AAV38499-Y38944). The antigenic proteins, their
 CC fragments, their nucleic acids and antibodies are used for diagnosis,
 CC prevention (as vaccines) or treatment of Neisseria infections,
 CC such as meningitis, septicemia and gonorrhea. Both organisms
 CC are closely related. Fragments of the nucleic acids are useful
 CC as hybridisation probes and antisense reagents.
 XX
 SQ Sequence 903 BP; 207 A; 228 C; 257 G; 211 T; 0 other;
 Query Match 6.3%; Score 47.4; DB 20; Length 903;
 Best Local Similarity 43.6%; Pred. No. 5.9e-05;
 Matches 213; Conservative 0; Mismatches 276; Indels 0; Gaps 0;
 QY 19 GAAGAGATAAACTCGTCTTTCGGAGAGCTGCAAGGACATAGCCACATAGT 78
 DB 130 GCAAGTGTGCTGCTTGGCACTGGTGGCGGCATTAAGAGATTGGCCATATAGAT 189
 QY 79 TTGAAGCTATTAACGAGCTCGTATTAAGGAGGCTTTAAGCGGGGTAGCGCCGG 138
 DB 190 GTTAAGGTTTGAAGAAAGAAACGGTATCTGTGAAGGTGTATCCGCAATCGGCA 249
 QY 139 GCAATCGTTCGCTTTTATGCTCAGGCTACCTCCCTGAAGGATGTTCAGCTTCTG 198
 DB 250 TCGATATGTCGGAGGCTTTTGGCATCGGTATGTGCGCCGACCGCTGAAATGGA 309
 QY 199 AAGAGGTAACCTGCTGAGAGCTGTTAAGTTCAAGCACTGTGAAGGATATAGG 258
 DB 310 GAGATTTTATAGTAACCGATTTAGTCATTTAATCTGTCCACAGAGTATTATCAA 369
 QY 259 TGGAGAAAGGCTATTAAGATTCTTGAGGAAGTCTCCCTTACAGAGATAGAAA 318
 DB 370 GCGAAGAAAGCTGCAAAATTAATCAACGAAAGTCGCGGAGGCAATTCAGAGTT 429
 QY 319 GAGATACGAGCTATATATGCGGAGGATTTATCTGGGAAGGCTCTATACCTTCG 378
 DB 430 CCCATCAATTTCCGCGCGTTCACATATTTTGAACCGGGAAGCGCTGCTTCAAT 489
 QY 379 GAAGGAGTTTAATCCCGCACTTCTCGGAGCTGTGCAATTCGCGCATATTTGA 438
 DB 490 CAAGGGAATCGCGGAGGCGGTGTGCTTCGCGGCAATTCGCAATGTTCAGCA 549
 QY 439 GTTGAATATAGAAATTAATCTGCTGTTACGAGGATATAGTTAACACTTCCTT 498
 DB 550 GTTCATCATCGGAGGACAAATATATGTGACGCGGTCTGTGCAACCGCTGCGCT 609
 QY 499 CCCTTCAG 507
 DB 610 GCCGCTCG 618
 RESULT 10
 AAV81946
 ID AAV81946 standard; DNA; 40138 BP.
 XX

AAV81946;
 19-OCT-1999 (first entry)
 V. marinus PKS-like cluster comprising ORFs 6,7,8 and 9.
 Polyketide-like synthesis; PKS; PKS-like gene; PUFA; DHA; transgenic; poly-unsaturated fatty acid; eicosapentenoic acid; docosahexanoic acid; EPA; oil; dietary supplement; infant feeding formulation; malnutrition; intravenous feeding formulation; cooking oil; fat; anti-inflammatory; cholesterol; open reading frame; ORF; ss.
 Vibrio marinus.

Key Location/Qualifiers
 CDS 17394..25352
 /tag= a
 /product= "ORF 6"
 CDS 25509..28160
 /tag= b
 /product= "ORF 7"
 CDS 28209..34265
 /tag= c
 /product= "ORF 8"
 CDS 34454..36118
 /tag= d
 /product= "ORF 9"

WO985625-A1.
 10-DEC-1998.
 04-JUN-1998; 98WO-US11639.
 04-JUN-1997; 97US-0048650.
 (CALJ) CALGENE LLC.
 Facciotti D, Lassner M, Metz JG;
 WPI; 1999-070271/06.
 New nucleic acid encoding polyketide-like synthesis enzymes of Vibrio marinus and transformed plants and microbes that produce polyunsaturated fatty acids, useful as pharmaceuticals and food supplements
 Example 1; Fig 5; 153pp; English.

The invention provides polyketide-like synthesis (PKS)-like genes that are used for the production of long chain polyunsaturated fatty acid (PUFA) productions. Genes responsible for eicosapentenoic acid (EPA) production in Shewanella putrefaciens and novel genes associated with the production of docosahexanoic acid (DHA) in Vibrio marinus are used to generate transgenic plants that can express transgenes encoding PKS-like genes associated with PUFA production. The PKS-like genes are used to transform plants and microbial cells to give recombinants having altered contents of PUFA (specifically DHA and EPA). Oils from these plants are useful as dietary supplements (in infant feeding formulations, to give a PUFA profile closer to that of human milk; for treating malnutrition; in intravenous feeding formulations; in cooking oils, fats etc.), also as anti-inflammatory agents and for reducing cholesterol levels. Fragments from the genes are useful as probes to isolate related molecules. or to detect organisms that express PKS-like genes. The method facilitates large scale production of PUFA by providing new pathways for their synthesis or suppressing interfering pathways. Expression of PUFA in seeds allows simple recovery, as oil which can be engineered to have a particular PUFA profile. Expression in microbes also allows simple recovery and control of PUFA profile and is not subject to external variables such as weather or food supply. The present sequence represents an approximately 40 kb PKS-like cluster DNA fragment from V. marinus comprising open reading frames (ORFs) 6, 7, 8 and 9.

SQ Sequence 40138 BP; 11845 A; 7635 C; 8758 G; 11900 T; 0 other;
 Query Match 6.1%; Score 46.2; DB 20; Length 40138;
 Best Local Similarity 52.9%; Pred. No. 0.0012;
 Matches 99; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
 QY 23 AGATAAACCTCGTTCTTTTCGGGAGGAGCTGCACAAAGGCATAGCCACATAGGTGTTTGA 82
 Db 14817 AAATTGGCTTAGTCTCTTCTGCGGTGGTGCAGAGGATTTGCTCATCTTGGTGATTAA 14876
 QY 83 AAGCTATAACAGCTCGGTATAGCGGTGAGGGCTTTAAGCGGGGTGAGCGCGGGGCAA 142
 Db 14877 AATACCTGTAGAGCAAGATATAAGACCGAATGTAATTGCGGGTACAAAGTGTGGCTCTA 14936
 QY 143 TCGTTTCGTCTTTATGCTCAGGCTACTCCCTGAAGGATGTTACGCCCTTCTGAAGA 202
 Db 14937 TGGTTGGTGCACCTTTATGCTCAGGACTTGAGATTGATGACATTTTACAAATCTTCATCG 14996
 QY 203 GGGTAAA 209
 Db 14997 ATGTAAA 15003

RESULT 11
 AAA711520
 ID AAA71520 standard; DNA; 40138 BP.
 AC AAA71520;
 XX
 DT 11-DEC-2000 (first entry)
 DE V. marinus PKS-like gene cluster encoding ORF6 to ORF9.
 KW PKS pathway; polyunsaturated long chain fatty acid; plant; transgenic;
 KW polyketide-like synthesis; PUFA; dietary supplement; intravenous feeding;
 KW malnutrition; cooking oil; cooking fat; margarine;
 KW docosahexenoic acid production; eicosapentenoic acid production; ds.
 XX
 OS Vibrio marinus.
 XX
 PN WO200042195-A2.
 XX
 PD 20-JUL-2000.
 XX
 PF 14-JAN-2000; 2000WO-US00956.
 XX
 PR 14-JAN-1999; 99US-0231899.
 XX
 PA (CALJ) CALGENE LLC.
 XX
 PI Facciotti D, Metz JG, Lassner M;
 XX
 DR WPI; 2000-476063/41.
 XX
 PT New DNA sequences encoding for polyketide (PK)-like synthesis pathway
 PT genes from Shewanella, Vibrio and Schizochtrium, useful for creating
 PT transgenic plants that express poly-unsaturated long chain fatty acids
 XX
 PS Example 1; Fig 5; 302pp; English.
 XX
 CC This invention describes novel DNA sequences encoding for polyketide
 CC (PK)-like synthesis (PKS-like) pathway genes from Shewanella, Vibrio
 CC and Schizochtrium. The nucleic acids are useful for isolating related
 CC molecules or in methods to detect organisms expressing the PKS-like
 CC genes. They are also useful for creating transgenic plants that express
 CC poly-unsaturated long chain fatty acids. The poly-unsaturated long chain
 CC fatty acids produced recombinantly are useful as dietary supplements for
 CC patients undergoing intravenous feeding or for preventing or treating
 CC malnutrition. The poly-unsaturated long chain fatty acids can also be
 CC incorporated into cooking oils, fats or margarine formulated so that in
 CC normal use the recipient receives a desired amount of poly-unsaturated
 CC long chain fatty acids. The nucleic acids are also useful in large scale

CC production of docosahexaenoic acid and eicosapentaenoic acid, and for the
CC modification of the fatty acid profile of host cells and edible plant
CC tissues and/or plant parts. Transgenic production of polyunsaturated
CC fatty acids in particular host cells allows quicker purification from
CC natural sources such as fish or plants. This sequence represents a
CC Vibrio marinus DNA fragment encoding the PKS gene cluster incorporating
CC ORF6 to ORF9.

XX
XX
XX Sequence 40138 BP; 11845 A; 7634 C; 8759 G; 11900 T; 0 other;

QY Query Match 6.1%; Score 46.2; DB 21; Length 40138;
Best Local Similarity 52.9%; Pred. No. 0.0012;
Matches 99; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 23 AGATTAACCTGCTTCTTGGGAGGAGCTGCAAGGCGCTTTAAGCGGGGTGAGCGCCGGGCA 82

DB 14817 AAATGGCTTAGTCTTCTTGGGAGGAGCTGCAAGGCGCTTTAAGCGGGGTGAGCGCCGGGCA 14876

QY 83 AAGCTATTAACGAGCTGCTGATATTAAGGCTTTAAGCGGGGTGAGCGCCGGGCA 142

DB 14877 AATACCTTTAGAGCAAGATTAAGACCGATTAATGCGGCTTACAAAGTCTGCTTA 14936

QY 143 TCGTTTCGCTTTTATGCTTCAAGCTACTCCCTGAAGGAGTTCAGGCTTCTGAAGA 202

DB 14937 TGGTTGGTCACTTATTTGCTCAGACTTGAGATTAATGATTTTACAATTTCTCATCG 14996

QY 203 GGGTAA 209

DB 14997 ATGTAA 15003

RESULT 12

AA164984
ID AA164984 standard; DNA; 41587 BP.

XX AA164984;

XX 04-DEC-2001 (first entry)

DE Moritella marina icosapentaenoic acid biosynthesis enzyme DNA #1.

XX Icosapentaenoic acid biosynthesis; docosahexaenoic acid productivity;

XX DHA industrial production; ds.

OS Moritella marina.

PN JP2001169780-A.

PD 26-JUN-2001.

PF 15-DEC-1999; 99JP-0356614.

PR 15-DEC-1999; 99JP-0356614.

PA (KEIZ-) KEIZAI SANGYOSHIO SANGYO GIJUTSU SOGO KEN.

XX WPI; 2001-592536/67.

PT Gene of a docosahexaenoic acid-producing microbe, used for industrial

XX production of docosahexaenoic acid -

XX Claim 4; Page 10-24; 68pp; Japanese.

XX The present invention provides a DNA sequence encoding an icosapentaenoic
CC acid-biosynthesizing enzyme group-like protein group derived from a
CC microbe having docosahexaenoic acid (DHA) productivity. The gene can be
CC used for the industrial production of DHA. The present sequence is a
CC coding sequence described in the exemplification of the invention.

XX Sequence 41587 BP; 12323 A; 7885 C; 9126 G; 12253 T; 0 other;

QY Query Match 6.1%; Score 46.2; DB 22; Length 41587;
Best Local Similarity 52.9%; Pred. No. 0.0012;

Matches 99; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 23 AGATTAACCTGCTTCTTGGGAGGAGCTGCAAGGCGCTTTAAGCGGGGTGAGCGCCGGGCA 82

DB 7433 AAATGGCTTAGTCTTCTTGGGAGGAGCTGCAAGGCGCTTTAAGCGGGGTGAGCGCCGGGCA 7492

QY 83 AAGCTATTAACGAGCTGCTGATATTAAGGCTTTAAGCGGGGTGAGCGCCGGGCA 142

DB 7493 AATACCTTTAGAGCAAGATTAAGACCGATTAATGCGGCTTACAAAGTCTGCTTA 7552

QY 143 TCGTTTCGCTTTTATGCTTCAAGCTACTCCCTGAAGGAGTTCAGGCTTCTGAAGA 202

DB 7553 TGGTTGGTCACTTATTTGCTCAGACTTGAGATTAATGATTTTACAATTTCTCATCG 7612

QY 203 GGGTAA 209

DB 7613 ATGTAA 7619

RESULT 13

ABK79457
ID ABK79457 standard; DNA; 467 BP.

XX ABK79457;

DE 13-AUG-2002 (first entry)

DE Bacillus clausii genomic sequence tag (GST) #2300.

KW Differential gene expression; genomic sequenced tag; GST;

KW altered culture condition; environmental stress;

KW physiological provocation; ds.

OS Bacillus clausii.

PN WO200229113-A2.

PD 11-APR-2002.

PF 05-OCT-2001; 2001WO-US31437.

PR 06-OCT-2000; 2000US-0680598.

PR 27-MAR-2001; 2001US-279526P.

PA (NOVO) NOVOZYMES BIOTECH INC.

PA (NOVO) NOVOZYMES AS.

PI Berka R, Clausen IG;

XX WPI; 2002-416684/44.

PT Monitoring differential expression of several genes in first Bacillus

XX cell relative to expression of same genes in one or more second

XX Bacillus cells, by using substrate containing Bacillus genomic

XX sequenced tag array -

XX Claim 11; SEQ ID NO 6748; 200pp; English.

XX The invention describes a method of monitoring differential expression of
CC genes in a first Bacillus cell relative to expression of the genes in
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC isolated from Bacillus cells to a substrate containing array of Bacillus
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first Bacillus cell relative to expression of the same genes
CC in one or more second Bacillus cells. The method is useful for monitoring
CC global expression of several genes from a Bacillus cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which Bacillus cells adapt to changes in culture conditions.
CC environmental stress or other physiological provocation. Extensive

CC follow-up characterisation is unnecessary, when one spot on an array
CC equals one gene or one open reading frame, since sequence information is
CC available. This sequence represents a genomic sequence tag (GST) used in
CC the method of the invention.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at

CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 467 BP; 122 A; 88 C; 133 G; 124 T; 0 other;

Query Match 6.1%; Score 45.8; DB 24; Length 467;

Best Local Similarity 44.4%; Pred. No. 0.00014;

Matches 185; Conservative 0; Mismatches 232; Indels 0; Gaps 0;

QY 24 GATTAACCTCGTCTTTCGGAGAGCTGCAAGGGCATAGCCACATAGGTGTTTGA 83

DB 21 GATTGGACTGGCCCTTGGGCTCTGGCGGGCTAGAGGCTTTGCCACATTTGGTTTAA 80

QY 84 AGCTATAACGAGCTCGGTATAGGGTGAAGGGCTTTAAGCGGGGTGAGCGCGGCAAT 143

DB 81 GACGCTTACAGAGCGGGGTTACATTTAGCTGTTAGCGAGCATGGTGCAAT 140

QY 144 CGTTTCGCTCTTTATGCTCAGGCTACTCCCTGAAGGGATGTTTCAAGCTTTCTGAAGAG 203

DB 141 GGTGGCGACTAIGTATGCGTAGGCCATTGCGTTGAAACGATGGAAGTTTTCGGAACA 200

QY 204 GGTAACTGGCTGAGCTGTTTAAAGTTCAGCCACCTCTGAGGGGATGATAGGGTGGGA 263

DB 201 TTTCAAAACGAAATATTTACTTTGGATTTTACAGTCTCAAAACAAAGGGCTTTTGTGACA 260

QY 264 GAAGCTATAAGATTCTCTGAGGAAGTTCTCCCTTACAGGAGATAGAAAACTTGAGAT 323

DB 261 AAGATTGAAAGCTCATTTGCTTTACTAGCTTAAAGACAAAGCTGGAGTCGCTTTTTC 320

QY 324 ACCGACGATATATGCGCGAGGATTTATCTCGGGAAGGGCTCTATACCTCTCGGAAGG 383

DB 321 GCCAGTTCAAGTGTAGCACTGATCTTTAAGTGGCCAAAAGTGTGTATGTCAAAGG 380

QY 384 GAGTTTAAATCCCGCACTTCTCGGACGCTGCAATTTCCCGGCAATTTTGAACCCGT 440

DB 381 AGACGTCGCCAAGCGGTTTCAGCAAGCTTTGCGATTTCCAGGCAATTTTGTCTCTGT 437

RESULT 14

AAZ12212

ID AAZ12212 standard; DNA; 447 BP.

AC AAZ12212;

DT 08-OCT-1999 (first entry)

DE Neisseria meningitidis partial ORF137 sequence.

NE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

NE treatment; Neisseria infection; meningitis; septicaemia; gonorrhea; ss.

OS Neisseria meningitidis.

XX WO9924578-A2.

XX 20-MAY-1999.

XX 09-OCT-1998; 98WO-IB01665.

XX 01-SEP-1998; 98GB-0019016.

XX 06-NOV-1997; 97GB-0023516.

XX 14-NOV-1997; 97GB-0024190.

XX 18-NOV-1997; 97GB-0024386.

XX 27-NOV-1997; 97GB-0025158.

XX 10-DEC-1997; 97GB-0026147.

XX 14-JAN-1998; 98GB-0000759.

PA (CHIR-) CHIRON SPA.

PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;

XX WPI; 1999-327407/27.

DR P-PSDB; AAY38777.

XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for

PT diagnosis, treatment and prevention of infection

XX Claim 9; Page 322; 524pp; English.

XX Nucleotide sequences AAZ11972-Z12358 represent open reading frames

CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode

CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their

CC fragments, their nucleic acids and antibodies are used for diagnosis,

CC prevention (as vaccines) or treatment of Neisseria infections,

CC such as meningitis, septicaemia and gonorrhea. Both organisms

CC are closely related. Fragments of the nucleic acids are useful

CC as hybridisation probes and antisense reagents.

XX SQ Sequence 447 BP; 110 A; 115 C; 114 G; 107 T; 1 other;

Query Match 5.6%; Score 42.6; DB 20; Length 447;

Best Local Similarity 54.0%; Pred. No. 0.0016;

Matches 87; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 19 GAAGAGATAAACCTCGTCTTTTCGGAGGAGCTGCAAGGGCATAGCCCATAGGTGTT 78

DB 130 GCAGTGGTGGTTCGGCACTCGGTGGCGGCGCATCTAAAGGATTTGCCCATGTAGTATT 189

QY 79 TTGAAGCTATAAACGAGCTGGTATAGGGTGAGGCTTTAAGCGGGTGAGCCCGGG 138

DB 190 ATTAAGGTTTGAAGAGAAACGGTATTCCTGTGAAGTGGTTACCGGCACCTCCGCGAGT 249

QY 139 GCAATCGTTTCGTTCTTTTATGCTCAGCTACTCCCTGA 179

DB 250 TCGATTGTCGGCAACCTTTTTCATCGGGTATGTCGCCCGA 290

RESULT 15

AAA81390

ID AAA81390 standard; DNA; 447 BP.

AC AAA81390;

DT 04-DEC-2000 (first entry)

DE N. meningitidis MenB polynucleotide sequence ORF number 66.

NE Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;

NE antigen; vaccine; diagnosis; infection; antibacterial; identification;

NE Meningococcus B; MenB; ds.

OS Neisseria meningitidis.

XX WO200022430-A2.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-US23573.

XX 09-OCT-1998; 98US-0103794.

XX 30-APR-1999; 99US-0132068.

XX (CHIR) CHIRON CORP.

XX Frazer CM, Hickey E, Peterson J, Tetzelin H, Venter JC;

XX Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;

XX PI Rappuoli R, Pizza M;

XX DR WPI; 2000-318079/27.

Thu Jun 19 17:30:36 2003

us-09-903-410-26.rml

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 13:34:10 ; Search time 47 Seconds

(without alignments)
4932.928 Million cell updates/sec

Title: US-09-903-410-26

Perfect score: 756
1 ttcgagatcgaggaattga.....tagagttacttcggaatag 756

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/1/ina/5A_COMB.seq: *
4: /cgn2_6/prodata/1/ina/5B_COMB.seq: *
5: /cgn2_6/prodata/1/ina/5C_COMB.seq: *
6: /cgn2_6/prodata/1/ina/5D_COMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	756	100.0	756	2	US-08-602-359A-26
2	397.8	52.6	750	2	US-08-602-359A-29
3	93.4	12.4	2108	4	US-09-221-017B-305
4	46.2	6.1	40138	3	US-09-090-793-12
5	40.4	5.3	7218	1	US-08-232-463-14
6	32.2	4.3	4403765	4	US-09-103-840A-2
7	32.2	4.3	4411529	4	US-09-103-840A-1
8	31.4	4.2	2057	3	US-09-008-303-1
9	31	4.1	2409	4	US-09-484-970B-101
10	30.4	4.0	1555	3	US-08-747-574-3
11	29.8	3.9	1867	2	US-08-607-509-3
12	29.8	3.9	1867	2	US-08-634-642-3
13	29.8	3.9	1867	2	US-08-989-370-3
14	29.6	3.9	756	2	US-08-602-359A-26
15	29.6	3.9	6735	4	US-08-561-527-104
16	29	3.8	14602	1	US-08-597-236-1
17	29	3.8	14602	1	US-08-597-236-1
18	28.8	3.8	10763	1	US-08-746-682A-1
19	28.8	3.8	10763	1	US-08-761-258-1
20	28.8	3.8	4403765	4	US-09-103-840A-2
21	28.8	3.8	4411529	4	US-09-103-840A-1
22	28.6	3.8	1803	4	US-08-976-259-71
23	28.6	3.8	5577	1	US-08-326-117B-1
24	28.6	3.8	5577	1	US-08-982-129-1
25	28.6	3.8	5582	4	US-09-178-176B-1
26	28.6	3.8	5582	4	US-09-457-864-1
27	28.4	3.8	510	4	US-09-328-111-722

C 28	28.4	3.8	1344	1	US-07-940-852-1	Sequence 1, Appli
C 29	28.4	3.8	1344	1	US-08-396-126-1	Sequence 1, Appli
C 30	28.4	3.8	1710	3	US-09-000-630C-1	Sequence 1, Appli
C 31	28.4	3.8	1710	3	US-08-862-730C-1	Sequence 1, Appli
C 32	28.2	3.7	951	1	US-08-671-525B-1	Sequence 1, Appli
C 33	28.2	3.7	951	1	US-08-672-109B-1	Sequence 1, Appli
C 34	28.2	3.7	951	1	US-08-842-045-1	Sequence 1, Appli
C 35	28.2	3.7	951	2	US-08-842-238-1	Sequence 1, Appli
C 36	28.2	3.7	951	2	US-08-629-335B-1	Sequence 1, Appli
C 37	28.2	3.7	1270	4	US-08-387-805-1	Sequence 1, Appli
C 38	28.2	3.7	1633	1	US-07-866-979-5	Sequence 5, Appli
C 39	28.2	3.7	1633	2	US-08-466-906B-5	Sequence 5, Appli
C 40	28.2	3.7	1633	3	US-08-706-281A-5	Sequence 5, Appli
C 41	28.2	3.7	1633	4	US-09-201-746-5	Sequence 5, Appli
C 42	28.2	3.7	1633	4	US-09-097-231-5	Sequence 5, Appli
C 43	28.2	3.7	1839	1	US-08-075-193-1	Sequence 1, Appli
C 44	28.2	3.7	1839	2	US-08-564-090A-1	Sequence 1, Appli
C 45	28.2	3.7	1839	5	PCT-US94-06698-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-602-359A-26
Sequence 26, Application US/08602359A
Patent No. 5942430
GENERAL INFORMATION:
APPLICANT: ROBERTSON, Daniel E.
APPLICANT: MURPHY, Dennis
APPLICANT: REID, John
APPLICANT: MARFIA, Anthony
APPLICANT: LINK, Steven
APPLICANT: SWANSON, Ronald V.
APPLICANT: WARREN, Patrick V.
APPLICANT: KOSMOTKA, Anna
TITLE OF INVENTION: ESTERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSER: FISH & RICHARDSON P. C.
STREET: 4225 EXECUTIVE SQUARE, STE 1400
CITY: LA JOLLA
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/602,359A
FILING DATE: February 16, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: HAILE, LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
TELEFAX: 619-678-5099
INFORMATION FOR SEQ ID NO: 26:
LENGTH: 756 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: GENOMIC DNA
US-08-602-359A-26

Query Match 100.0%; Score 756; DB 2; Length 756;
 Best Local Similarity 100.0%; Pred. No. 3.7e-241;
 Matches 756; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGAGTTGAGGAATTTGAGAGATAAAGCTCGTTCTTCGGAGAGCTGCAAGGCG 60
 Db 1 TTGAGTTGAGGAATTTGAGAGATAAAGCTCGTTCTTCGGAGAGCTGCAAGGCG 60

Qy 61 ATAGCCACATAGGTTGTTGAAAGCTATAAAGAGCTCGGTATAAGGCTGAGGCTTTA 120
 Db 61 ATAGCCACATAGGTTGTTGAAAGCTATAAAGAGCTCGGTATAAGGCTGAGGCTTTA 120

Qy 121 AGCGGGTGAGCGCGGGGCAATCGTTTCGGTCTTTATGCTCAGGCTACTCCCTGAA 180
 Db 121 AGCGGGTGAGCGCGGGGCAATCGTTTCGGTCTTTATGCTCAGGCTACTCCCTGAA 180

Qy 181 GGGATGTTGAGCTTCTGAAAGAGGTTAACTGGCTGAAGCTGTTTAAAGTTCAAGCCACT 240
 Db 181 GGGATGTTGAGCTTCTGAAAGAGGTTAACTGGCTGAAGCTGTTTAAAGTTCAAGCCACT 240

Qy 241 CTGAGAGTATGATAGGTTGAGAGGCTATAAGATTCCTGAGAGGTTCTCCCTTAC 300
 Db 241 CTGAGAGTATGATAGGTTGAGAGGCTATAAGATTCCTGAGAGGTTCTCCCTTAC 300

Qy 301 AGGAGATAGAAAATTTGAGATACCGAGCTATATATGCGGACGAGTTTATCTCGGGA 360
 Db 301 AGGAGATAGAAAATTTGAGATACCGAGCTATATATGCGGACGAGTTTATCTCGGGA 360

Qy 361 AGGGCTCTATACCTCTCGGAGGAGTTTAACTCCCGCACTCTCGGAGCTGTCGCAAT 420
 Db 361 AGGGCTCTATACCTCTCGGAGGAGTTTAACTCCCGCACTCTCGGAGCTGTCGCAAT 420

Qy 421 CCGGCAATTTGAAACCGGTTGAGTATAAGAAATTTACTTGCTCGTTGACGAGGTATAGTT 480
 Db 421 CCGGCAATTTGAAACCGGTTGAGTATAAGAAATTTACTTGCTCGTTGACGAGGTATAGTT 480

Qy 481 AACACCTTCGCTGAGGCTTTTACGAAAGGAGTTTAACTCCCGCACTCTCGGAGCTGTCGCAAT 540
 Db 481 AACACCTTCGCTGAGGCTTTTACGAAAGGAGTTTAACTCCCGCACTCTCGGAGCTGTCGCAAT 540

Qy 541 CTTCCATAGAGCGGAAAGGATATAAAGAACTTTTCAATCTCTTTCGAGGAGCTTC 600
 Db 541 CTTCCATAGAGCGGAAAGGATATAAAGAACTTTTCAATCTCTTTCGAGGAGCTTC 600

Qy 601 TTTCTGCGGTCGCTCAAACTCCGAAAGAGGAGGTTTGTGCACTCGTTATAGTT 660
 Db 601 TTTCTGCGGTCGCTCAAACTCCGAAAGAGGAGGTTTGTGCACTCGTTATAGTT 660

Qy 661 CTTGAGCTGAGGAGTTTCAACCCCTTGTATGATAAAGCGGACCAAAATATGAGAGG 720
 Db 661 CTTGAGCTGAGGAGTTTCAACCCCTTGTATGATAAAGCGGACCAAAATATGAGAGG 720

Qy 721 GCATACATAAAGGCTTAGAGGTACTTTCTGATAG 756
 Db 721 GGTATACATAAAGGCTTAGAGGTACTTTCTGATAG 756

RESULT 2
 US-08-602-359A-29
 ; Sequence 29, Application US/08602359A
 ; Patent No. 5942430
 ; GENERAL INFORMATION:
 ; APPLICANT: ROBERTSON, Daniel E.
 ; APPLICANT: MORPHY, Dennis
 ; APPLICANT: REID, John
 ; APPLICANT: MAFFIA, Anthony
 ; APPLICANT: LINK, Steven
 ; APPLICANT: SWANSON, Ronald V.
 ; APPLICANT: WARREN, Patrick V.
 ; APPLICANT: KOSMOTKA, Anna
 ; TITLE OF INVENTION: ESTERASES
 ; NUMBER OF SEQUENCES: 42
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: FISH & RICHARDSON P.C.
 STREET: 4225 EXECUTIVE SQUARE, STE 1400
 CITY: LA JOLLA
 STATE: CALIFORNIA
 COUNTRY: USA
 ZIP: 92037
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 INCH DISKETTE
 COMPUTER: IBM PS/2
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: WORD PERFECT 6.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/602,359A
 FILING DATE: February 16, 1996
 CLASSIFICATION: 435
 PRIOR APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: HAILE, LISA A.
 REGISTRATION NUMBER: 38,347
 REFERENCE/DOCKET NUMBER: 09010/010001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619-678-5070
 TELEFAX: 619-678-5099
 INFORMATION FOR SEQ ID NO: 29:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 750 NUCLEOTIDES
 TYPE: NUCLEIC ACID
 STRANDEDNESS: SINGLE
 TOPOLOGY: LINEAR
 MOLECULE TYPE: GENOMIC DNA
 US-08-602-359A-29

Query Match 52.6%; Score 397.8; DB 2; Length 750;
 Best Local Similarity 72.4%; Pred. No. 3.2e-122;
 Matches 516; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

Qy 1 TTGAGTTGAGGAATTTGAGAGATAAAGCTCGTTCTTCGGAGAGCTGCAAGGCG 60
 Db 13 TTGAAATTTGAGAGGTTTGAAGAGTTAACTTAGTTCTTTCGGAGGCGCTGCCAAGGT 72

Qy 61 ATAGCCACATAGGTTGTTGAAAGCTATAAAGAGCTCGGTATAAGGCTGAGGCTTTA 120
 Db 73 ATGCCCATATAGGTTTAAAGCTCTGAGAGCTCGGTATAAGGTTAAAGAGCTC 132

Qy 121 AGCGGGTGAGCGCGGGGCAATCGTTTCGGTCTTTTATGCTCAGGCTACTCCCTGAA 180
 Db 133 AGCGGGTTAAGTCTGAGAGCTATCGTTTTCGGTCTTTTACGCTTCGGGCTACATCCCGAC 192

Qy 181 GGGATGTTGAGCTTCTGAAAGAGGTTAACTGGCTGAAAGCTTTTAAAGTTCAAGCCACT 240
 Db 193 GAGATGTTAAAGCTCTGAAAGAGCTAACTGGCTCAAACTTTTAAAGTTCAAAACACCG 252

Qy 241 CTGAGAGGATGATAGGTTGAGAGGCTATAAGATTCCTTGAGAGGTTCTCCCTTAC 300
 Db 253 AAAATGGGCTTAAATGGGTTGGGAGGCTGCAAGTTTTCGAAAAAGAGCTCGAGATT 312

Qy 301 AGGAGAAATAGAAAACCTTGAGATACCGAGCTATATATCGCGAGCGGATTTTACTCGGGA 360
 Db 313 AAGAGGCTGGAAGACCTGAACATACCAACCTATCTTCCTCGCGGATCTGTACACCGGA 372

Qy 361 AGGGCTCTATACCTCTCGGAGGAGTTTAACTCCCGCACTCTCGGAGCTGTCGCAAT 420
 Db 373 AAGGCTCTTACTTTCGGAGAGGTTGACTTAAATTCCTGCTTCGGAGGTTGTTCCATA 432

Qy 421 CCGGCAATTTGAAACCGGTTGAGTATAAGAAATTTACTTGCTCGTTGACGAGGTATAGTT 480
 Db 433 CCGGCAATTTTGAACCAAGTTCAGTACGAGAAATTTTCTACTTTTGAACGAGGTATAGTG 492

Qy 481 AACACCTTCGCTGAGGCTTTTACGAAAGGAGGTTTTCGCAAGGTTTGGTGTGATGTC 540
 Db 493 AACACCTTCGCTGAGGCTTTTGAAGAGGTTTCAAGAAACCCATAATTCGGGGTAGATGTG 552

Page 3

US-09-221-017B-305

```

; SOFTWARE: PatentIn Ver. 2.0

```

```

; SEQ ID NO 12
; LENGTH: 40138
; TYPE: DNA
; ORGANISM: Vibrio marinus
; US-09-090-793-12

Query Match      6.1%; Score 46.2; DB 3; Length 40138;
Best Local Similarity 52.9%; Pred. No. 0.00016;
Matches 99; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 23 AGATAAAGCTGCTTCTTCGGGAGAGCTGCAAGGGGATAGCCACATAGGCTTTTGA 82
Db 14817 AAATGCGCTAGTCTCTTCGGGAGAGCTGCAAGGGGATAGGCTTTTGA 14876

QY 83 AGCTATAAAGAGCTGCGTATAGGCTGAGGCTTTAAGCGGGTGAAGCGCGGCA 142
Db 14877 AATACCTGTAGAGCAAGATATAGCAAGATGATTTGCGGGTACAGTCTGCTCTA 14936

QY 143 TGGTTTGGCTTTTATGCTCAGGCTACTCCCTGGAAGGGATGTTGAGCCTTCTGA 202
Db 14937 TGGTTTGGCTCACTTTATGCTCAGGCTGAGGATGATGACATTTTACATCTTCATCG 14996

QY 203 GGGTAAA.209
Db 14997 ATGTAAA 15003

```

```

RESULT 5
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKING NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:

```

```

; CLONE: PTZgpt-Fls
; US-08-232-463-14

Query Match      5.3%; Score 40.4; DB 1; Length 7218;
Best Local Similarity 7.4%; Pred. No. 0.0052;
Matches 32; Conservative 206; Mismatches 192; Indels 0; Gaps 0;

QY 3 GAGATTGAGGAAATTTGAGAGATATAACCTCGTCTTCTTCGGGAGGAGCTGCAAGGCGCAT 62
Db 1454 GAGATAGAGAAATTTGAGAGATATAACCTCGTCTTCTTCGGGAGGAGCTGCAAGGCGCAT 1395

QY 63 AGCCACATAGGTTTGTAAAGCTATAACAGAGCTCGGTATAGGAGGAGGCTTTAAG 122
Db 1394 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1335

QY 123 CGGGGTAGCGCGGGGCAATCGTTTCGGTCTTTTATGCTCAGGCTACTCCCTGAAGG 182
Db 1334 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1275

QY 183 CATGTTTCAGCCTTCTGAAGAGGGTAAACTCGCTGAGCTGTTTAAAGTTCAAGCCACTCT 242
Db 1274 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1215

QY 243 GAAGGATTGATAGGTCGGGAGAGGCTATAAGATTCTTTCAGGAAGTTCTCCCTACAG 302
Db 1214 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1155

QY 303 GAGATAGAAAACTTGAGATACCGACTATATATGCGCGAGCTTATATCTCGGGAAG 362
Db 1154 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1095

QY 363 GGCTCTATACCTCTCGGAGGAGTTTAAATCCCGCAGCTTCTCGGAGCTGTGCAATCC 422
Db 1094 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1035

QY 423 CGGCATATTT 432
Db 1034 GGAATTAAT 1025

```

```

RESULT 6
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
; US-09-103-840A-2

```

```

Query Match      4.3%; Score 32.2; DB 4; Length 4403765;
Best Local Similarity 52.6%; Pred. No. 52;
Matches 70; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 404 TCGCGAGCTGTGCAATTCGCCGATATTTGAACCGCTTGATAGATATAAGAAATTTACTGTGCG 463
Db 370685 TCGCGGAGTTGAGAGAGCGGTGTGTGTGCTGCGGTGTTTCCAGCCGCGAGTTTGGC 370744

```

QY 464 TTGACGAGGTATAGTAAACACTTCCCGTGGAGCCCTTTCAGAAAGCGTATTCGCA 523
DB 370745 GTGCGCGGTGTTCGCGATGCAACGTTTCGTTGCCGAGTTGAAAGCCGATTTTCCG 370804
QY 524 CCGTTTGCGTTGA 536
DB 370805 CTGCCCGAGTTGA 370817

RESULT 7
US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 4.3%; Score 32.2; DB 4; Length 4411529;
Best Local Similarity 52.6%; Pred. No. 52;
Matches 70; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 404 TCGGAGCGTGTGCATTCCTCCGACATTTGAAACCGTTGAGTATTAAGATTACTGCTCG 463
DB 370628 TCGCCGAGTTAAAGAGCGGTGTGTGCTGCGCGGTTTCCAGCCCGAGATTGCG 370687
QY 464 TTGACGAGGTATAGTAAACACTTCCCGTGGAGCCCTTTCAGAAAGCGTATTCGCA 523
DB 370688 GTGCGCGGTGTTCGCGATGCAACGTTTCGTTGCCGAGTTGAAAGCCGATTTTCCG 370747
QY 524 CCGTTTGCGTTGA 536
DB 370748 CTGCCCGAGTTGA 370760

RESULT 8
US-09-008-303-1/C
Sequence 1, Application US/09008303
Patent No. 6033889
GENERAL INFORMATION:
APPLICANT: HAN, YE SUN
APPLICANT: YU, GYU YU
APPLICANT: KIM, SUNG HOU
APPLICANT: LIM, JAE HWAN
APPLICANT: RYU, JAE RYON
APPLICANT: CHOI, IN GEOL
TITLE OF INVENTION: GENE SEQUENCE OF AQUIFEX PYROPHILUS
TITLE OF INVENTION: SUPEROXIDE DISMUTASE AND PROTEIN EXPRESSED IN ESCHERICHIA
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: OHLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/008,303
FILING DATE: 16-JAN-1998
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 97-1140
FILING DATE: 16-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: OHLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2901-0109-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-1000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2057 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1014..1652
US-09-008-303-1

Query Match 4.2%; Score 31.4; DB 3; Length 2057;
Best Local Similarity 54.9%; Pred. No. 2.5;
Matches 62; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 433 GAACCGGTGTAGTAAATTAATCTGCTGTTGACGAGGTATTAACACTTCCG 492
DB 541 GAAGCCTTGAAGAGCAATATTAAGACCGTTGCGAGCTATTCCTGAAACTTGAG 482
QY 493 GTGAGCCCTTTCAGAAAGCGTATTCACCGTTTGCCTTATCTCTCC 545
DB 481 GTAATGAGAGAGGAGGAGGTAATAATACCGTTAGCTTGAATGCTCC 429

RESULT 9
US-09-484-970B-101/C
Sequence 101, Application US/09484970B
Patent No. 6426186
GENERAL INFORMATION:
APPLICANT: Jones, Karen A.
APPLICANT: Volkmuth, Wayne
APPLICANT: Walker, Michael G.
TITLE OF INVENTION: BONE REMODELING GENES
FILE REFERENCE: PB-0014 US
CURRENT APPLICATION NUMBER: US/09/484,970B
CURRENT FILING DATE: 2000-01-18
NUMBER OF SEQ ID NOS: 172
SOFTWARE: PERL Program
SEQ ID NO 101
LENGTH: 2409
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6426186 404040.2CBI
NAME/KEY: unsure
LOCATION: 11-13, 15-17, 25, 35, 1273, 1281, 1288, 2402
OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-101

Query Match 4.1%; Score 31; DB 4; Length 2409;
Best Local Similarity 64.8%; Pred. No. 3.7;
Matches 46; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 46 GGAGCTGCAAGGCGATAGCCCATAGTGTGTTGAAGCTATATAACGAGCTCGGTATA 105
Db 876 GCACGACAAAGCGATTTCACATATGTTTGGAGGTAATTAAGTAACCTCTGTATA 817
QY 106 AGGCTGAGGCG 116
Db 816 AAAATAAATGC 806

RESULT 10

US-08-747-574-3/c
; Sequence 3, Application US/08747574
; Patent No. 6015939
; GENERAL INFORMATION:
; APPLICANT: CALGENE, INC.
; TITLE OF INVENTION: PLANT VDE GENES AND
; TITLE OF INVENTION: METHODS RELATED THERETO
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/747,574
; FILING DATE: No. 6015939ember 7, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/023,502
; FILING DATE: August 6, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/006,315
; FILING DATE: No. 6015939ember 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: 119-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1555
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA

US-08-747-574-3

Query Match 4.0%; Score 30.4; DB 3; Length 1555;
Best Local Similarity 53.3%; Pred. No. 4.6;
Matches 64; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 582 CATCTTTGAGGAGCTTCTTTCTGCGGTCGAGTCAAACTCGAAAGAAAGAGGTT 641
Db 311 CAGCTCTTTCAGCTCAATCTTCGATGTAATGGCAATGTCAATATCCCTTTAGAGAA 252
QY 642 TTGTGACCTCGTTATAGTCTCTGAGCTTGAGGAGTTACACACCTTGTATGTTAGAAAGC 701
Db 251 TCTTGACCTGAATCCAGATAAAGACGTGAGGATCTCCACCAGTTGTTCTTGAGATCRGC 192

RESULT 11

US-08-607-509-3
; Sequence 3, Application US/08607509
; Patent No. 5876735
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: METHODS FOR ENHANCEMENT OF PROTECTIVE IMMUNE RESPONSES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/607,509
; FILING DATE: 16-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.404C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1867 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 117..1325

US-08-607-509-3

Query Match 3.9%; Score 29.8; DB 2; Length 1867;
Best Local Similarity 48.5%; Pred. No. 8.1;
Matches 82; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
QY 126 GGTGAGCGCGGGCAATCGTTTCGGTCTTTTATGCTCTCAGGCTACTCCCTCGAAGGAT 185
Db 641 GGTGCTCGACGAGGCTGATGATGCTCTCAGGGCTTCGCGACCAAGATTTACGAGAT 700
QY 186 GTTCAGCCTTCTGAAAGGGTAAACTGGCTGAAGCTGTTTAAGTTCAAGCCACTCTGAA 245
Db 701 CTTCGGCTTCTCGCGAAGGACATCCAGGTCGGCTCTTCTCCGCCACGATGCCGGAGGA 760
QY 246 GGGATTGATAGGTTGGGAGAGGCTATAAGATTCTTTGAGGAAGTTCTC 294
Db 761 GGTACTGGAGCTGACGAGAAGTTCTATCGCGCACCCCGTGCGTATTCTC 809

RESULT 12

US-08-634-642-3
; Sequence 3, Application US/08634642
; Patent No. 5879687
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: METHODS FOR ENHANCEMENT OF
; TITLE OF INVENTION: PROTECTIVE IMMUNE RESPONSES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA

Page 7

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1325
US-08-989-370-3

Query Match 3.9%; Score 29.8; DB 3; Length 1867;
Best Local Similarity 48.5%; Pred. No. 8.1;
Matches 82; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 126 GGTGAGGCGCGGGGCATCGTTTCGGCTTTTATGCTAGCGCTACCTCCCTGAAGGAT 185
|||
Db 641 GGTCTCGACGAGGCTGATAGATGCTGTCTGAGGGCTTCGCGACCGAATTAACGAGAT 700

Qy 186 GTTCAGCGCTTCTGAAGAGGTTAACTGGCTGAAGCTTTTAAGTTCAAGCACCTCGAA 245
|||
Db 701 CTTCGCTTCTTCGCCAGAGACATTCAGGTGCGGCTCTTTCGCCACGATGCCGAGGA 760

Qy 246 GGAATGATAGGGGTGGAGAGGCTATTAAGATTCCTTGAGGAAGTTCTC 294
|||
Db 761 GGATCTGAGCTGACGAGGAAGTTCAATGCGCGACCCCGTGTATTTCTC 809

RESULT 14
US-08-602-359A-26/c
Sequence 26, Application US//08602359A
Patent No. 5942430
GENERAL INFORMATION:
APPLICANT: ROBERTSON, Daniel E.
APPLICANT: MURPHY, Dennis
APPLICANT: REID, John
APPLICANT: MAFFIA, Anthony
APPLICANT: LINK, Steven
APPLICANT: SWANSON, Ronald V.
APPLICANT: WARREN, Patrick V.
APPLICANT: KOSMOTKA, Anna
TITLE OF INVENTION: ESTERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & RICHARDSON P.C.
STREET: 4225 EXECUTIVE SQUARE, STE 1400
CITY: LA JOLLA
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/602.359A
FILING DATE: February 16, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: HAILE, LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
TELEFAX: 619-678-5099
INFORMATION FOR SEQ ID NO: 26:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 756 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: GENOMIC DNA
US-08-602-359A-26

Query Match 3.9%; Score 29.6; DB 2; Length 756;
Best Local Similarity 53.4%; Pred. No. 5.7;
Matches 62; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
QY 516 TATTCCACCGTTTCGGTGTGATGCTTCCATAGACCGGAAAGGATATAAGAACAT 575
DB 631 TCTTTTCGGAGTTTGAGCGGACCGCAAGAAAGAGCTCTCAAAAGGATGTGAAGATGT 572
QY 576 TCTTCACATCTTTTGGAGGAGCTTCTTTCTTGGGGTCCGCTCAAACTCCGAAAAGA 631
DB 571 TCTTTATATCTTTTCCGGCTCTATGGGAAGGACATCAAGCAACGGTGGGAATA 516

RESULT 15

US-08-961-527-104
Sequence 104, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 6735 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-104

Query Match 3.9%; Score 29.6; DB 4; Length 6735;
Best Local Similarity 59.5%; Pred. No. 19;
Matches 50; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
QY 193 CTTCTGAAGAGGGTAAACTGGCTGTAAGCTGTTTAAGTTCAGCCACCTCTGAAGGGATTG 252
DB 5464 CTTTTCGATATGCTAAACTTTGTGAGACGACGACGAGTTAAGGCCATCTATGAGACTATT 5523
QY 253 ATAGGGTGGGAGAGGCTATAGA 276

Db 5524 TAAATGAATGGCAAGTTACTAAGA 5547
Search completed: June 19, 2003, 14:22:18
Job time : 64 secs

Thu Jun 19 17:30:37 2003

us-09-903-410-26.rmpb

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 14:02:55 ; Search time 118 Seconds

(without alignments)
9401.483 Million cell updates/sec

Title: US-09-903-410-26

Perfect score: 756
Sequence: 1 ttgagattgaggaatttga.....tagagttacttctgatatag 756

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Published Applications NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	756	100.0	756	9	US-10-027-805-26
2	756	100.0	756	9	US-10-027-804-26
3	756	100.0	756	10	US-09-903-410-26
4	397.8	52.6	750	9	US-10-027-805-29
5	397.8	52.6	750	9	US-10-027-804-29
6	397.8	52.6	750	10	US-09-903-410-29
7	46.2	6.1	40138	9	US-10-331-061-12
8	45.8	6.1	467	10	US-09-974-300-6748
9	35.8	4.7	876	10	US-09-974-300-2350
10	35	4.6	1062	9	US-09-938-842A-1639
11	35	4.6	1390	9	US-10-295-403-89
12	33.2	4.4	458	9	US-10-184-644-546
13	33.2	4.4	458	9	US-10-184-634-546
14	33.2	4.4	5520	9	US-09-902-525-38
15	32.8	4.3	2000	10	US-09-887-576-66
16	32.8	4.3	2004	10	US-09-887-576-184
17	32.6	4.3	494	10	US-09-864-761-14137
18	32.4	4.3	505	10	US-09-783-590-5019
19	32.2	4.3	286	10	US-09-983-965-4221

20	32	4.2	1332	9	US-09-938-842A-1753	Sequence 1753, Ap
21	32	4.2	17569	9	US-09-736-457-1804	Sequence 1804, Ap
22	32	4.2	17569	9	US-09-902-941-1804	Sequence 1804, Ap
23	32	4.2	17569	9	US-09-849-626-1804	Sequence 1804, Ap
24	32	4.2	17569	9	US-10-017-754-1804	Sequence 1804, Ap
25	31.8	4.2	476	10	US-09-864-761-14137	Sequence 14137, A
26	31.6	4.2	378361	9	US-09-901-136-3	Sequence 3, Appli
27	31.6	4.2	1503841	9	US-09-946-807-1	Sequence 1, Appli
28	31.6	4.2	1503841	10	US-09-795-668-1	Sequence 1, Appli
29	31.6	4.2	1503841	10	US-09-795-668-1	Sequence 1, Appli
30	31.4	4.2	7004	9	US-10-239-676-119	Sequence 119, Ap
31	31	4.1	653	9	US-10-184-644-402	Sequence 402, Ap
32	31	4.1	653	9	US-10-184-644-402	Sequence 402, Ap
33	31	4.1	1622	9	US-09-822-846-407	Sequence 407, Ap
34	31	4.1	4138	10	US-09-802-472B-5	Sequence 5, Appli
35	30.8	4.1	11928	9	US-10-199-024-55	Sequence 55, Appli
36	30.8	4.1	11928	9	US-10-199-024-56	Sequence 56, Appli
37	30.8	4.1	11928	9	US-10-199-024-57	Sequence 57, Appli
38	30.8	4.1	11928	9	US-10-199-024-58	Sequence 58, Appli
39	30.8	4.1	684973	10	US-09-263-959-1	Sequence 1, Appli
40	30.6	4.0	442	9	US-09-933-797-266	Sequence 266, Ap
41	30.6	4.0	657	12	US-10-003-356-3	Sequence 3, Appli
42	30.6	4.0	751	10	US-09-974-300-1332	Sequence 1332, Ap
43	30.6	4.0	2781	12	US-10-003-356-9	Sequence 9, Appli
44	30.6	4.0	368004	10	US-09-949-654-3	Sequence 3, Appli
45	30.4	4.0	407	10	US-09-960-352-2555	Sequence 2555, Ap

ALIGNMENTS

RESULT 1
US-10-027-805-26
Sequence 26, Application US/10027805
Patent No. US20020164725A1
GENERAL INFORMATION:
APPLICANT: ROBERTSON, Daniel E.
MURPHY, Dennis
REID, John
MAFFIA, Anthony
LINK, Steven
SWANSON, Ronald V.
WARREN, Patrick V.
KOSMOTKA, Anna
TITLE OF INVENTION: ESTERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSER: FISH & RICHARDSON P.C.
STREET: 4225 EXECUTIVE SQUARE, STE 1400
CITY: LA JOLLA
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/027,805
FILING DATE: 21-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/602,359
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: HALL, LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
TELEFAX: 619-678-5099
INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:
LENGTH: 756 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: GENOMIC DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 26;
US-10-027-805-26

Query Match 100.0%; Score 756; DB 9; Length 756;
Best Local Similarity 100.0%; Pred. No. 1.9e-247;
Matches 756; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 TTGAGATTGAGGAAATTTGAAGAGATAAACCCTGTTCTTCGCGAGAGCTGCAAGGGC 60
Db 1 TTGAGATTGAGGAAATTTGAAGAGATAAACCCTGTTCTTCGCGAGAGCTGCAAGGGC 60
Qy 61 ATAGCCACATAGTGTGTTTGAAGAGCTATAAAGAGCTCGGTATAGGGTGAAGGGCTTTA 120
Db 61 ATAGCCACATAGTGTGTTTGAAGAGCTATAAAGAGCTCGGTATAGGGTGAAGGGCTTTA 120
Qy 121 AGCGGGGTGAGCGCGGGGCAATCGTTTGGTCTTTTATGCTTCAGGCTACTCCCTGAA 180
Db 121 AGCGGGGTGAGCGCGGGGCAATCGTTTGGTCTTTTATGCTTCAGGCTACTCCCTGAA 180
Qy 181 GCGATGTTCAAGCTTCTGAAGAGGTAATACTGGCTGAAGCTGTTTAAAGTTCAAGCACT 240
Db 181 GCGATGTTCAAGCTTCTGAAGAGGTAATACTGGCTGAAGCTGTTTAAAGTTCAAGCACT 240
Qy 241 CTGAAGGATTTGATAGGGTGGGAGAGGCTATAAGATTCCTTGAGGAAGTTCTCCCTTAC 300
Db 241 CTGAAGGATTTGATAGGGTGGGAGAGGCTATAAGATTCCTTGAGGAAGTTCTCCCTTAC 300
Qy 301 AGGAGATAGAAAACCTTGAGATACCGAGCTATATATGCGGACGGAATTTACTCGGGA 360
Db 301 AGGAGATAGAAAACCTTGAGATACCGAGCTATATATGCGGACGGAATTTACTCGGGA 360
Qy 361 AGGGCTCTATACCTCTCGGAAGGGAGTTTAACTCCCGCACTTCTCGGAGCTGTCGAAT 420
Db 361 AGGGCTCTATACCTCTCGGAAGGGAGTTTAACTCCCGCACTTCTCGGAGCTGTCGAAT 420
Qy 421 CCGGCAATTTGAAACCGGTTGAGTATAAGAAATTAATCTGCTCGTTGACGGAGTATAGT 480
Db 421 CCGGCAATTTGAAACCGGTTGAGTATAAGAAATTAATCTGCTCGTTGACGGAGTATAGT 480
Qy 481 AACAACTTCCTCGGTTGAGCTTTCAGGAAGGGATTTCCACCGCTTTCGCTGATGTC 540
Db 481 AACAACTTCCTCGGTTGAGCTTTCAGGAAGGGATTTCCACCGCTTTCGCTGATGTC 540
Qy 541 CTTCCCAATAGAGCGGAAAGGATATAAGAAACATTCCTCACATCCTTTTGAGGAGCTTC 600
Db 541 CTTCCCAATAGAGCGGAAAGGATATAAGAAACATTCCTCACATCCTTTTGAGGAGCTTC 600
Qy 601 TTTCTTGGCTCGCTCAAACTCCGAAAGAGAAAGAGTTTGTGACTCGTTATAGT 660
Db 601 TTTCTTGGCTCGCTCAAACTCCGAAAGAGAAAGAGTTTGTGACTCGTTATAGT 660
Qy 661 CTTGAGCTTGAGGAGTTTCAACCCCTTGATGTTAGAAAGCGGCAATTAATCGGAGG 720
Db 661 CTTGAGCTTGAGGAGTTTCAACCCCTTGATGTTAGAAAGCGGCAATTAATCGGAGG 720
Qy 721 GGATACATAAAGGCTTAGAGGTACTTTCTGAATAG 756
Db 721 GGATACATAAAGGCTTAGAGGTACTTTCTGAATAG 756
```

RESULT 2
US-10-027-804-26
Sequence 26, Application US/10027804
Publication No. US20030054530A1
GENERAL INFORMATION:
APPLICANT: ROBERTSON, Daniel E.
MURPHY, Dennis

REID, John
MAFFIA, Anthony
LINK, Steven
SWANSON, Ronald V.
WARREN, Patrick V.
KOSMOTKA, Anna
TITLE OF INVENTION: ESTERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSSEE: FISH & RICHARDSON P.C.
STREET: 4225 EXECUTIVE SQUARE, STE 1400
CITY: LA JOLLA
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/027,804
FILING DATE: 21-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/602,359
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: HAILE, LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
TELEFAX: 619-678-5099
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 756 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: GENOMIC DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 26;
US-10-027-804-26

Query Match 100.0%; Score 756; DB 9; Length 756;
Best Local Similarity 100.0%; Pred. No. 1.9e-247;
Matches 756; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 TTGAGATTGAGGAAATTTGAAGAGATAAACCCTGTTCTTCGCGAGAGCTGCAAGGGC 60
Db 1 TTGAGATTGAGGAAATTTGAAGAGATAAACCCTGTTCTTCGCGAGAGCTGCAAGGGC 60
Qy 61 ATAGCCACATAGTGTGTTTGAAGAGCTATAAAGAGCTCGGTATAGGGTGAAGGGCTTTA 120
Db 61 ATAGCCACATAGTGTGTTTGAAGAGCTATAAAGAGCTCGGTATAGGGTGAAGGGCTTTA 120
Qy 121 AGCGGGGTGAGCGCGGGGCAATCGTTTGGTCTTTTATGCTTCAGGCTACTCCCTGAA 180
Db 121 AGCGGGGTGAGCGCGGGGCAATCGTTTGGTCTTTTATGCTTCAGGCTACTCCCTGAA 180
Qy 181 GCGATGTTCAAGCTTCTGAAGAGGTAATACTGGCTGAAGCTGTTTAAAGTTCAAGCACT 240
Db 181 GCGATGTTCAAGCTTCTGAAGAGGTAATACTGGCTGAAGCTGTTTAAAGTTCAAGCACT 240
Qy 241 CTGAAGGATTTGATAGGGTGGGAGAGGCTATAAGATTCCTTGAGGAAGTTCTCCCTTAC 300
Db 241 CTGAAGGATTTGATAGGGTGGGAGAGGCTATAAGATTCCTTGAGGAAGTTCTCCCTTAC 300
Qy 301 AGGAGATAGAAAACCTTGAGATACCGAGCTATATATGCGGACGGAATTTACTCGGGA 360
Db 301 AGGAGATAGAAAACCTTGAGATACCGAGCTATATATGCGGACGGAATTTACTCGGGA 360
Qy 361 AGGGCTCTATACCTCTCGGAAGGGAGTTTAACTCCCGCACTTCTCGGAGCTGTCGAAT 420
```

Db 361 AGGGCTCTATACCTCTCGGAGAGATTAAATCCCGCACTTCGCGAGCTGCAATT 420
Qy 421 CCGGCAATATTGAACCGGTTGATATAGAAATTAATCTGCTGTTGACGAGATAGT 480
Db 421 CCGGCAATATTGAACCGGTTGATATAGAAATTAATCTGCTGTTGACGAGATAGT 480
Qy 481 AACAACTTCCGTTGAGCCCTTTCAGAAAGCGGATTCACCGTTTGCGTTGATGTC 540
Db 481 AACAACTTCCGTTGAGCCCTTTCAGAAAGCGGATTCACCGTTTGCGTTGATGTC 540
Qy 541 CTTCCTATAGAGCGGAAAGATATTAAGAAACATCTTCAATCTCTTGAAGAGATTC 600
Db 541 CTTCCTATAGAGCGGAAAGATATTAAGAAACATCTTCAATCTCTTGAAGAGATTC 600
Qy 601 TTTCTTGCGGTCGCTCAAACTCCGAAAGAGAGATTGTTGACCTGTTATAGT 660
Db 601 TTTCTTGCGGTCGCTCAAACTCCGAAAGAGAGATTGTTGACCTGTTATAGT 660
Qy 661 CCGGCAATATTGAACCGGTTGATATAGAAATTAATCTGCTGTTGACGAGATAGT 720
Db 661 CCGGCAATATTGAACCGGTTGATATAGAAATTAATCTGCTGTTGACGAGATAGT 720
Qy 721 GGATACATTAAGGCTTGAAGGTAATCTTCTGAATAG 756
Db 721 GGATACATTAAGGCTTGAAGGTAATCTTCTGAATAG 756

RESULT 3

US-09-903-410-26
Sequence 26, Application US/09903410
Patent No. US20020146799A1
GENERAL INFORMATION:
APPLICANT: DIVERSEA CORPORATION
APPLICANT: ROBERTSON, Dan
APPLICANT: MURPHY, Dennis
APPLICANT: REID, John
APPLICANT: MAFELA, Anthony
APPLICANT: LINK, Steven
APPLICANT: SWANSON, Ronald
APPLICANT: WARREN, Patrick
APPLICANT: KOSMOTKA, Anna
TITLE OF INVENTION: ENZYMES HAVING ESTERASE ACTIVITY AND METHODS OF USE THEREOF
FILE REFERENCE: DIVERSEA-180-2
CURRENT APPLICATION NUMBER: US/09/903,410
PRIOR FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: US 09/382,242
PRIOR FILING DATE: 1999-08-24
PRIOR APPLICATION NUMBER: US 08/602,359
PRIOR FILING DATE: 1996-02-16
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn version 3.0
SEQ ID NO 26
LENGTH: 756
TYPE: DNA
ORGANISM: Aquifex pyrophilus
US-09-903-410-26

Query Match 100.0%; Score 756; DB 10; Length 756;
Best Local Similarity 100.0%; Ered. No. 1,9e-247;
Matches 756; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTGAGATGAGGAAATTTGAAGATAAACCTCGTTCTTGGGAGAGAGCTGCAAGAGGC 60
Db 1 TTGAGATGAGGAAATTTGAAGATAAACCTCGTTCTTGGGAGAGAGCTGCAAGAGGC 60
Qy 61 ATAGCCACATAGTGTGTTTGAAGCTATAACGAGCTCGGTATTAAGGGTGAAGGCTTTA 120
Db 61 ATAGCCACATAGTGTGTTTGAAGCTATAACGAGCTCGGTATTAAGGGTGAAGGCTTTA 120
Qy 121 AGCGGGGTAGAGCGCGGAGCATGTTTCCGCTTTTATGCTCAGCTACTCCCGTGA 180
Db 121 AGCGGGGTAGAGCGCGGAGCATGTTTCCGCTTTTATGCTCAGCTACTCCCGTGA 180

Qy 181 GGGATGTCACGCTTCTGAAGAGGTTAACTGCGTGAAGCTGTTTATGTTCAAGCACC 240
Db 181 GGGATGTCACGCTTCTGAAGAGGTTAACTGCGTGAAGCTGTTTATGTTCAAGCACC 240
Qy 241 CTGAAGGATTAAGTGGTGGAGAGAGCTATTAAGATTCCTTGAAGAGTTCCTCTTAC 300
Db 241 CTGAAGGATTAAGTGGTGGAGAGAGCTATTAAGATTCCTTGAAGAGTTCCTCTTAC 300
Qy 301 AGGAGATTAAGAAACTTGAATACGACGTATATATAGCGGAGATTTATCTCGGGA 360
Db 301 AGGAGATTAAGAAACTTGAATACGACGTATATATAGCGGAGATTTATCTCGGGA 360
Qy 361 AGGCTCTATACCTCTCGGAGAGGATTATCCCGCACTTCTCGGAGCTGTCATTT 420
Db 361 AGGCTCTATACCTCTCGGAGAGGATTATCCCGCACTTCTCGGAGCTGTCATTT 420
Qy 421 CCGGCAATATTGAACCGGTTGATATAGAAATTAATCTGCTGTTGACGAGATAGT 480
Db 421 CCGGCAATATTGAACCGGTTGATATAGAAATTAATCTGCTGTTGACGAGATAGT 480
Qy 481 AACAACTTCCGTTGAGCCCTTTCAGAAAGCGGATTCACCGTTTGCGTTGATGTC 540
Db 481 AACAACTTCCGTTGAGCCCTTTCAGAAAGCGGATTCACCGTTTGCGTTGATGTC 540
Qy 541 CTTCCTATAGAGCGGAAAGATATTAAGAAACATCTTCAATCTCTTGAAGAGATTC 600
Db 541 CTTCCTATAGAGCGGAAAGATATTAAGAAACATCTTCAATCTCTTGAAGAGATTC 600
Qy 601 TTTCTTGCGGTCGCTCAAACTCCGAAAGAGAGATTGTTGACCTGTTATAGT 660
Db 601 TTTCTTGCGGTCGCTCAAACTCCGAAAGAGAGATTGTTGACCTGTTATAGT 660
Qy 661 CCGGCAATATTGAACCGGTTGATATAGAAATTAATCTGCTGTTGACGAGATAGT 720
Db 661 CCGGCAATATTGAACCGGTTGATATAGAAATTAATCTGCTGTTGACGAGATAGT 720
Qy 721 GGATACATTAAGGCTTGAAGGTAATCTTCTGAATAG 756
Db 721 GGATACATTAAGGCTTGAAGGTAATCTTCTGAATAG 756

RESULT 4

US-10-027-805-29
Sequence 29, Application US/10027805
Patent No. US20020164725A1
GENERAL INFORMATION:
APPLICANT: ROBERTSON, Daniel E.
APPLICANT: MURPHY, Dennis
APPLICANT: REID, John
APPLICANT: MAFELA, Anthony
APPLICANT: LINK, Steven
APPLICANT: SWANSON, Ronald V.
APPLICANT: WARREN, Patrick V.
APPLICANT: KOSMOTKA, Anna
TITLE OF INVENTION: ESTERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESSES:
ADDRESS: FISH & RICHARDSON P.C.
STREET: 4225 EXECUTIVE SQUARE, STE 1400
CITY: LA JOLLA
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/027,805
FILING DATE: 21-Dec-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/602,359
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: HAILE, LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
TELEFAX: 619-678-5099
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 750 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: GENOMIC DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 29:
US-10-027-805-29

Query Match 52.6%; Score 397.8; DB 9; Length 750;
Best Local Similarity 72.4%; Pred. No. 4e-125;
Matches 516; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

QY 1 TTGAGATTCAGCAATTTGAGAGATAAAGCTGCTTCTTCGGAGAGAGCTGCAAGGCG 60
Db |||||
QY 61 ATAGCCCATATAGTGTGTTTGAAGCTATAAAGCTCGGTATAGGGGTGAGGCTTTA 120
Db |||||
QY 73 ATCGCCCATATAGTGTGTTTGAAGCTATAAAGCTCGGTATAGGGGTGAGGCTTTA 132
Db |||||
QY 121 AGCGGGTTCAGCGCGGCGCAATCGTTCGCTCTTTATGCTCAGGCTACTCCCTCGAA 180
Db |||||
QY 133 AGCGGGTTCAGCGCGGCGCAATCGTTCGCTCTTTATGCTCAGGCTACTCCCTCGAA 192
Db |||||
QY 181 GGATGTTTCAGGCTTCGAGAGGTTAACTGGCTGAGGCTGTTTAAAGTTCAGGCTACT 240
Db |||||
QY 193 GAGATGTTAAACTCTGTAAGAGGTAAGCTGAGCTCAAACTTTTAAAGTTCAGGCTACT 252
Db |||||
QY 241 CTGAAGGATTCATAGGTTGGAGAGGCTATAAGATTCCTTGAAGAGGTTCTCCCTTAC 300
Db |||||
QY 253 AAAATGGGCTTAATGGGTTGGAGAGGCTCGAGATTTTGGAAAAGAGCTCGAGTT 312
Db |||||
QY 301 AGGAGATGAAATCTTGGATACCGAGCTATATATATATATATATATATATATATATAT 360
Db |||||
QY 313 AAGAGCTGGAAGCTGAACTACCACTTCTTCTCGGCGGATCTGTACAGGGA 372
Db |||||
QY 361 AGGCTCTATATCTTCGGAAGGAGTTTAAATCCCGGCTTCTCGGAGCTGTGCAATT 420
Db |||||
QY 373 AAGGCTCTTATCTTCGGAAGGAGTTTAAATCCCGGCTTCTCGGAGCTGTGCAATT 432
Db |||||
QY 421 CCGGCAATTTGAAACCGCTTCAGTATAGAAATTTCTTCTGTTGAGGAGGATAGTT 480
Db |||||
QY 433 CCGGCAATTTGAAACCGCTTCAGTATAGAAATTTCTTCTGTTGAGGAGGATAGTT 492
Db |||||
QY 481 AACACCTTCGCTTCAGGCTTCAGGAAAGCGGTATTCACCGCTTTCGCTGATATGTC 540
Db |||||
QY 493 AACACCTTCGCTTCAGGCTTCAGGAAAGCGGTATTCACCGCTTTCGCTGATATGTC 552
Db |||||
QY 541 CTTCCCATAGAGCGGGAAGGATATAAGAACTCTTCACTCTTTTGAAGGAGCTTC 600
Db |||||
QY 553 CTTCCCATAGAGCGGGAAGGATATAAGAACTCTTCACTCTTTTGAAGGAGCTTC 612
Db |||||
QY 601 TTTCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCT 660
Db |||||
QY 613 TTTCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCT 672
Db |||||
QY 661 CTTGAGCTTCGAGGTTCAACCCCTTCGATGTTTGAAGAGCGGACCAATAAT 713
Db |||||
QY 673 CCTCCCTTCGAGAGTTCT 725
Db |||||

RESULT 5
US-10-027-804-29
Sequence 29, Application US/10027804
Publication No. US20030054530A1
GENERAL INFORMATION:
APPLICANT: ROBERTSON, Daniel E.
MURPHY, Dennis
REID, John
MAFFIA, Anthony
LINK, Steven
SWANSON, Ronald V.
WARREN, Patrick V.
KOSMOTKA, Anna
TITLE OF INVENTION: ESTERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & RICHARDSON P. C.
STREET: 4225 EXECUTIVE SQUARE, STE 1400
CITY: LA JOLLA
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/027,804
FILING DATE: 21-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/602,359
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: HAILE, LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5099
TELEFAX: 619-678-5099
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 750 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: GENOMIC DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 29:
US-10-027-804-29

Query Match 52.6%; Score 397.8; DB 9; Length 750;
Best Local Similarity 72.4%; Pred. No. 4e-125;
Matches 516; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

QY 1 TTGAGATTCAGCAATTTGAGAGATAAAGCTGCTTCTTCGGAGAGAGCTGCAAGGCG 60
Db |||||
QY 13 TTGAAATTCAGAGGTTTGAAGAGGTTAACTAGTTCTTTCGGAGGCGCTGCAAGGTT 72
Db |||||
QY 61 ATAGCCCATATAGTGTGTTTGAAGCTATAAAGCTCGGTATAGGGGTGAGGCTTTA 120
Db |||||
QY 73 ATCGCCCATATAGTGTGTTTGAAGCTATAAAGCTCGGTATAGGGGTGAGGCTTC 132
Db |||||
QY 121 AGCGGGTTCAGCGCGGCGCAATCGTTCGCTCTTTATGCTCAGGCTACTCCCTCGAA 180
Db |||||
QY 133 AGCGGGTTCAGCGCGGCGCAATCGTTCGCTCTTTATGCTCAGGCTACTCCCTCGAA 192
Db |||||
QY 181 GGATGTTTCAGGCTTCGAGAGGTTAACTGGCTGAGGCTGTTTAAAGTTCAGGCTACT 240
Db |||||
QY 193 GAGATGTTAAACTCTGTAAGAGGTAAGCTGAGCTCAAACTTTTAAAGTTCAGGCTACT 252
Db |||||
QY 241 CTGAAGGATTCATAGGTTGGAGAGGCTATAAGATTCCTTGAAGAGGTTCTCCCTTAC 300
Db |||||

Thu Jun 19 17:30:37 2003

us-09-903-410-26.rnpb

Page 5

Db	253	AAATGGGCTTAAATGGGCGTGGAGAGCGTGAAGATTTTTGGAAAAAGAGCTGGAGTT	312
Qy	301	AGGAGAAATAGAAAAACTTTGAGATACCGAGGTATATATGCGCGAGGAGTTTATCTCGGGA	360
Db	313	AAGAGGCTGGAAACCTTGAAACATCCMACCCATCTTGTCTCGCGGAGTCTGTACACGGGA	372
Qy	361	AGGAGCTTATACCTCTCGAGAGGGAGTTTAATCCCGACATCTTCGGCAGCTGNGCAATT	420
Db	373	AAGGCTTTTACTTCGCGAGAGGTGACTTAATTTCCCGTCTTCGAGAGTTGTTCATA	432
Qy	421	CCGGGCAATTTTGAACCCCGTTGAGTAAAGATTATCTGCTGTGACGAGGTATAGTT	480
Db	433	CCGGGGAATTTTAAACAGTTGAGTACGGAATTTTCTATCTGTGTGACGAGGTATAGTG	492
Qy	481	AACAACTTCCCGTTGAGGCGCTTTTCAGAAACGGGTATTCGACCGTTGCTGTGANGTC	540
Db	493	AACAACTCGCGGTAGAACCTTGTGGAAAAATTCAAGAAACCCATATATCGGGTATGAATGTG	552
Qy	541	CTTCCCATAGAGCCGGAAAAAGATATTAAGAACATTTCTTCACATCTTTTGGAGAGCTTC	600
Db	553	CTTCCCATTAATCTGAAGAAAGAAAGATTAAAAATATATCTCCACATCTTTATAGGAGCTTC	612
Qy	601	TTTCTTCCGCGTCCGCTCAAACTCCAAAAAGAAAGAGTTTGTGACCTCGTTATAGTT	660
Db	613	TTTCTCGCGGTTGTTCCTCAATTGCAAAAAGAAAGAGAGTTCTGCAAGTAGTATAGAA	672
Qy	661	CTGAGAGTTGAGAGTTTACACCCCTGTAGTGTTAAGAAAAGCGGACCAAAATAT	713
Db	673	CTCTCCCTTGAAGAGTTCTCTCTCTGAGAGCTTAATTAAGCGGAGCAAGATATTT	725

RESULT 6

```

US-09-903-410-29
: Sequence 29, Application US/09903410
: Patent No. US20020146799A1
: GENERAL INFORMATION:
: APPLICANT: DIVERSA CORPORATION
: APPLICANT: ROBERTSON, Dan
: APPLICANT: MOREHY, Dennis
: APPLICANT: REID, John
: APPLICANT: MAFELA, Anthony
: APPLICANT: LINK, Steven
: APPLICANT: SWANSON, Ronald
: APPLICANT: WARREN, Patrick
: APPLICANT: KOSMOTKA, Anna
: TITLE OF INVENTION: ENZYMES HAVING ESTERASE ACTIVITY AND METHODS OF USE THEREOF
: FILE REFERENCE: DIVERS180-2
: CURRENT APPLICATION NUMBER: US/09/903,410
: CURRENT FILING DATE: 2001-07-10
: PRIOR APPLICATION NUMBER: US 09/382,242
: PRIOR FILING DATE: 1999-08-24
: PRIOR APPLICATION NUMBER: US 08/602,359
: PRIOR FILING DATE: 1996-02-16
: NUMBER OF SEQ ID NOS: 42
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 29
: LENGTH: 750
: TYPE: DNA
: ORGANISM: Aquifex VF5-34LC
: US-09-903-410-29

```

Query Match	52.6%;	Score 397.8;	DB 10;	Length 750;
Best Local Similarity	72.4%;	Pred. No. 4e-125;		
Matches	516;	Conservative	0;	Mismatches 197; Indels 0; Gaps 0

QY	1	TTNAGATGAGGAATTTGAAAGATTAACCGCTCTTTCGGAGAGAGCTGGAAGGGC	60
DB	13	TTGAATTTGAAGAGGTTTGAAGGCTTACCTTACTTCTTTCGGAGAGGCGTCGAAAGGCT	72
QY	61	ATAGCCCACTAAGCTGTTTTGAAGCTATAACGAGCTCGGTATTAAGGCTGAAGGCTTTA	120
DB	73	ATGCCCCATATAGAGCTTTTAAAGCTCTGGAAGAGCTCGGTATTAAGGCTGAAGAGCTTC	132

Qy	121	AGCGGGGTGAGAGCGCGGGGCATCGTTCCGTCCTTTTAAGTCCGACGGCTACCCCTGAA	180
Dp	133	AGCGGGGTAAAGCTGAGCTAGACTACGTTTCGGTCTTTTTCGCTTCGGAGCTACCTCCGAC	192
Qy	181	GGGATGTCAGCCCTTCGAAAGGGGTAAACTGCGTGAAGCTGTTTAAGTTCAGCCACT	240
Dp	193	GAGATGTTTAAACTCTCTGAAAGAGGTAAACGCGCTCAACTTTTCTTAAGTTCAAAACCG	252
Qy	241	CTGAAGGGATGTTAGGGAGGGAGAAAGGCTTAAGATTCCTTGAGGAAGTTCCTCCCTTAC	300
Dp	253	AAATATGGCTTTATGGGGTGGGAGAGAGCTGCAGATTTTGTGAAAAAGCTCGAGATT	312
Qy	301	AGAGATATGAAAACTTGAGATCCGACGTATATATGCGCAGCGATTATATCTCGGGA	360
Dp	313	AAGAGGCTGAAAGACTGAACATACCAACCTATCTTGTGCGCGGATCTGTACACGGGA	372
Qy	361	AGGGCTATATACCTTCGGAGGGAGTTTATTCGCCGCACTTCGCGGAGCTGTGCAATT	420
Dp	373	AAGGCTTTTACTTCGGCAGAGGTGACTTTATTCGCGTGCTTCGGAAGTTGTTCCATA	432
Qy	421	CCCGCATATTTGAACCCGTTGAGTATAGAAATTACTTGTCGTGTGAGCGAGGTATAGTT	480
Dp	433	CCCGGAGTTTTTAACCAAGTTATGATGACGAATTTTCTACTGTTGTGACGGAAGTATATG	492
Qy	481	AACAACCTTCCCGTGAAGCCTTTGAGAAAGCGGTATTCGCCACGTTTGCGCTGANTGC	540
Dp	493	AACAACCTGCGCGGTAGACCTTTGGAATAATTTCAAAAGAACCAATATCGGGGTAGATGTG	552
Qy	541	CTTCCCATTAAGCCGGAATAAGATATTAAGAACTTTTACATCTCTTTTGAAGACTTC	600
Dp	553	CTTCCCATTAAGCTAAGAAAGAAAGATTAATAATATCTCAACCTTATTAAGAGCTTC	612
Qy	601	TTTCTTGCGGTCCGCTCAACTCCGAAAGAAGAGGTTTGTGACCTCGTATAGTT	660
Dp	613	TTTCTTGCGGTTGTTCTCAATTCGAAAGAAGAAAGAGTTTCGACGCTGATTATGAA	672
Qy	661	CTGAGCGTTAGAGTTCAACCCCTGTATGTTGAATAAGCGGACCAATAT	713
Dp	673	CTTCCCTTGAAGATTCCTCTCTGAGCGTAAATTAAGCGGACGAGTAT	725

RESULT 7

```

US-10-331-061-12
: Sequence 12, Application US/10331061
: Publication No. US20030101486A1
: GENERAL INFORMATION:
: APPLICANT: Lasserer, Michael
: APPLICANT: Metz, James G
: APPLICANT: Facciotti, Daniel
: TITLE OF INVENTION: SCHIZOCHYTRIUM PKS GENES
: FILE REFERENCE: CGNE.131.0205
: CURRENT APPLICATION NUMBER: US/10/331,061
: CURRENT FILING DATE: 2002-12-27
: PRIOR APPLICATION NUMBER: US/09/331,899
: PRIOR FILING DATE: 1999-01-14
: PRIOR APPLICATION NUMBER: 60/048,650
: PRIOR FILING DATE: 1997-06-04
: PRIOR APPLICATION NUMBER: 09/090,793
: PRIOR FILING DATE: 1998-06-04
: NUMBER OF SEQ ID NOS: 86
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO. 12
: LENGTH: 40138
: TYPE: DNA
: ORGANISM: Vibrio marinus
US-10-331-061-12

```

Query Match	6.1%	Score 46.2;	DB 9;	Length 4038;
Best Local Similarity	52.9%;	Pred. No.	0.00045;	
Matches	99;	Mismatches	88;	Indels 0;
		Gaps	0;	
23	ACATTAACCTGTTTTCGGAGAGCTGCATAAGGCCATAGCCCAATTAGGTGTTTGA	82		

Db 14817 AAATTGCTTAGTCTCTTCTGGCGGTGTCGAAAGTAATGCTCATCTTGCTGATTAA 14876
Qy 83 AAGCTATATAACGAGCTCGGTATAGGCTTAAAGCGGGGTGAGCGCGGCGAA 142
Db 14877 AATACCTGTAGAGCAAGATATAGACGAATGTAATGCGGGTACAAGTGTGCTCTA 14936
Qy 143 TCGTTTCGCTCTTTATGCTCAGGCTACTCCCTGAGGATGTTTCAGCCCTTCTGAAGA 202
Db 14937 TGGTTGTGCACTTTATGCTCAGGACTTGAGATTGATGACATTTTACAAATCTTCATCG 14996
Qy 203 GGGTAAA 209
Db 14997 ATGTAAA 15003

RESULT 8

US-09-974-300-6748
; Sequence 6748, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Id Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6748
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Bacillus clausii
US-09-974-300-6748

Query Match 6.1%; Score 45.8; DB 10; Length 467;

Best Local Similarity 44.4%; Pred. No. 4.7e-05;

Matches 185; Conservative 0; Mismatches 232; Indels 0; Gaps 0;

Qy 24 GATTAACCTCGTCTTTTCGGGAGGCTGCAAAAGGCTAGCCACATAGGTGTTTGA 83
Db 21 GATTGATGCGCCCTTGGCTCTGCGGGCTAGAGGCTTTGCCACATTTGGTGTAA 80
Qy 84 AGCTATAACGAGCTCGGTATAAGGGTGAAGGCTTTAAAGCGGGTGAAGCGCGGCAAT 143
Db 81 GACGCTTACAGAAAGCGGGTTCACATTTAGCTGTAGCAGCATGGTGCAAT 140
Qy 144 CGTTTCGGTCTTTATGCTCAGGCTACTCCCTGAGGATGTTTCAGCTTCTGAAGAG 203
Db 141 GGTGGGCACTAATGAGCGTGAAGCCATTGGTTGAACCATGGAAGTTTTCGCAACA 200
Qy 204 GGTAACTGGCTGAAGCTTTTAAGTTCAAGCCACTCTGAAGGATGATAGGGTGA 263
Db 201 TTTCAACGAAATATTAATCTTGGATTTTACAGTCTCAAAACAAGGCTTATGCTGACA 260
Qy 264 GAAGGCTATAAGATCTTTCAGGAGTCTCCCTTACAGGAGATAGAAAATCTGAGAT 323
Db 261 AAGATTGAAGCTCATCTGTTTACTAGCTAAAGACAAAGCTGAGTGCCTTTTCC 320
Qy 324 ACGAGGTATATATCGGAGCGATTTATCTCCGGAAGGCTCTATACCTCTCGGAAG 383
Db 321 GCCAGTTCAAGTGGTAGCACTGATCTTTTAAGTGGCCAAAAGTGTGATGTCAAGGG 380
Qy 384 GAGTTTAATCCCGCACTTCTCGGAGCTGTGCAATCCCGGATATTTCAACCCGT 440
Db 381 AGACGTGCGCAAGCGGTTCGAGCAAGCTTGTGATTCAGGCATTTTGTCTCT 437

RESULT 9

US-09-974-300-2350
; Sequence 2350, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Id Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2350
; LENGTH: 876
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2350

Query Match 4.7%; Score 35.8; DB 10; Length 876;

Best Local Similarity 46.6%; Pred. No. 0.18;

Matches 115; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

Qy 23 AGATAAACCTCGTCTTTTCGGGAGGCTGCAAAAGGCTAGCCACATAGGTGTTTGA 82
Db 5 ACATCGACGGGTTTTTTTCAGCGCGCATGAAGGGTTGCGCTTGTGGAGCATACG 64
Qy 83 AAGCTATAAACGAGCTCGGTATAAGGGTGAAGGCTTTAAAGCGGGTGAAGCGCGGCA 142
Db 65 AAGCGCTTGAACAAAGCGGCTTCCGATTTAAAGGCTGCGCAACGAGCGCGTTTGA 124
Qy 143 TCGTTTCGCTCTTTATGCTCAGGCTACTCCCTGAGGATGTTTCAGCTTCTGAAGA 202
Db 125 TCATTCCTCATTTATCGCGCGCATATACAGCGGAGATCGGCAATGATGAGG 184
Qy 203 GGTAAACTGGCTGAAGCTTTTAAGTTCAAGCCACTCTGAAGGATGATAGGGTGG 262
Db 185 AATTAATGAAGCGAGCTCTTTCAGCGAGGTTTCACTCTTGCCTTTGAAATGCTGC 244
Qy 263 AGAAGGC 269
Db 245 AATGGGC 251

RESULT 10

US-09-938-842A-1639
; Sequence 1639, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRI1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1639
; LENGTH: 1062
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana

US-09-938-842A-1639

Query Match 4.6%; Score 35; DB 9; Length 1062;
Best Local Similarity 49.2%; Pred. No. 0.37;
Matches 92; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 196 CTGAAGAGGTAACCTGCTGAGCTTTTAAGTTACGCACTCTGAAGGATTTGATA 255
DB 71 CCGCGCGCGCAAAATGGAAGATACCGCTTGAAGAGAGCTGCTTGCAGGGATTCACG 130
QY 256 GGGTGGGAGAGGCTATAGATTCTTGAAGAGGATTCCTCCCTTACAGAGAAATGAAAA 315
DB 131 GGTGAGAGAGTTCTTAACTGATCGGTCAAGATCAACCAACGAGAGAGATCAGA 190
QY 316 CTGAGATACCGACGTATATATATGCGCGAGGATTTATATCTCGGAGAGGCTATATCTC 375
DB 191 CGAGATTAACCGCGGTGACTGACGTCCGCCGTTAACAGCTTCAAGAAAGTCAATTTCTCTAC 250
QY 376 TCGGAG 382
DB 251 TCGGTAG 257

RESULT 11

US-10-295-403-89
; Sequence 89, Application US/10295403
; Publication No. US20030101481A1

GENERAL INFORMATION:

APPLICANT: Heard, Jacqueline
APPLICANT: Riechmann, Jose Luis
APPLICANT: Adam, Luc
APPLICANT: Broun, Pierre
APPLICANT: Pineda, Omaire
APPLICANT: Reuber, Lynne
APPLICANT: Jiang, Cai-Zhong
APPLICANT: Keddie, James
APPLICANT: Zhang, James
APPLICANT: Benito, Maria-Ines
APPLICANT: Yu, Guo-Liang
APPLICANT: Fromm, Mike
TITLE OF INVENTION: PLANT GENE SEQUENCES I
FILE REFERENCE: M81-0003
CURRENT APPLICATION NUMBER: US/10/295,403
CURRENT FILING DATE: 2002-11-15
PRIOR APPLICATION NUMBER: US/09/394,519
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: 60/101,349
PRIOR FILING DATE: 1998-09-22
PRIOR APPLICATION NUMBER: 60/103,312
PRIOR FILING DATE: 1998-10-06
PRIOR APPLICATION NUMBER: 60/108,734
PRIOR FILING DATE: 1998-11-17
PRIOR APPLICATION NUMBER: 60/113,409
PRIOR FILING DATE: 1998-12-22
NUMBER OF SEQ ID NOS: 170
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 89
LENGTH: 1390
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: CDS
LOCATION: (35)..(1093)
OTHER INFORMATION: G1022
US-10-295-403-89

Query Match 4.6%; Score 35; DB 9; Length 1390;
Best Local Similarity 49.2%; Pred. No. 0.43;
Matches 92; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 196 CTGAAGAGGTAACCTGCTGAGCTTTTAAGTTACGCACTCTGAAGGATTTGATA 255
DB 105 CCGCGCGCGCAAAATGGAAGATACCGCTTGAAGAGAGCTGCTTGCAGGGATTCACG 164

QY 256 GGGTGGGAGAGGCTATAGATTCTTGAAGAGGATTCCTCCCTTACAGAGAAATGAAAA 315
DB 165 GTGAGAGAGGATTTCTTAACTGATCGGTCAAGATCAACCAACGAGAGAGATCAGA 224
QY 316 CTGAGATACCGAGTATATATATGCGGAGGATTTATATCTCGGAGAGGCTCTATATCTC 375
DB 225 CGAGATTAACCGCGGTGACTGACGTCCGCCGTTAACAGCTTCAAGAAAGTCAATTTCTCTAC 284
QY 376 TCGGAG 382
DB 285 TCGGTAG 291

RESULT 12

US-10-184-644-546
; Sequence 546, Application US/10184644
; Publication No. US20030044930A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Guiney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zhenli
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C227
CURRENT APPLICATION NUMBER: US/10/184,644
CURRENT FILING DATE: 2002-06-28
PRIOR APPLICATION removed - See file wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO: 546
LENGTH: 458
TYPE: PRT
ORGANISM: Homo Sapien
US-10-184-644-546

Query Match 4.4%; Score 33.2; DB 9; Length 458;
Best Local Similarity 9.9%; Pred. No. 0.94;
Matches 40; Conservative 113; Mismatches 252; Indels 0; Gaps 0;

QY 218 AGCTGTTAAGTTCAAGCCCTCTGAAGGATTTAGGTGGGAGGAGGCTATATAGT 277
DB 53 AVLGALYLSKSDRQAHATKCTSPSTBLSTSRATLLVATKASPGKLEBARALNQ 112
QY 278 TCCCTTGAAGAGTTCCTCCCTTACAGAGATGATAAATCTGAGTACCGATATATAT 337
DB 113 ALEMKRQKKEKQKLFMHALKMDPFDALTEFGIFSEBDKDIQADVLYTRALLISPY 172
QY 338 GCGGAGAGATTTATCTGCGGAGAGGCTCTATACCTCCGAGAGGATTTATCCCG 397
DB 173 HEKLVNRDRLPLVEIDRYSIISDKVKWMSIPKNSALRRMEETVYHHYHTVA 232
QY 398 CACTTCGCGAGCTGTCATATCCCGCATATTTAACCCTTGAGTATTAAGTAACT 457
DB 233 IEGNTLTLSEIRHLERAVVPKSLSEQNEVIGMHAAMKYINTLLVSIIGVTTSDVLE 292
QY 458 TGCCTGTTGACGAGATATGTTAAACACCTTCCCGTTAGCCCTTTCAGGAAGCGGTA 517
DB 293 IHRRLVGVDPVEAGRPRTTQVLVGHRIIPHPDVEKQMOEFVQWLSSEANLHVEFA 352
QY 518 TTCCACCGTTTGCTGATGTCCTTCCATAGAGCGGAGAAAGATATTAAGACATTC 577
DB 353 ALAHYKLVYIHPFDNGRTSRLIMULIMQAGYPTITRKGRSDYHYHLEAANBGDVR 412
QY 578 TTCAATCCTTTTGAAGAGCTTCTTTCTGCGGTCCGCTCAACT 622

Thu Jun 19 17:30:37 2003

us-09-903-410-26.rnpb

Page 9

CURRENT FILING DATE: 2001-06-25
PRIOR APPLICATION NUMBER: US 60/213,848
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 60/214,087
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 60/258,692
PRIOR FILING DATE: 2000-12-29
NUMBER OF SEQ ID NOS: 875
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 66
LENGTH: 2000
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-887-576-66

Query Match 4.3%; Score 32.8; DB 10; Length 2000;
Best Local Similarity 52.1%; Pred. No. 3;
Matches 73; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 35 TTCTTTCGGAGAGCTGCAAGGCAATAGGTCCTTTGAAAGCTATAACG 94
DB 188 TTTTGTATATATGATGATGATGATGATGATGATGATGATGATGAT 247
QY 95 AGCTGGTATAGGCTGAGGCTTTAAGCGGGTGAAGCCCGGGCAATGTTGGCT 154
DB 248 ATCAAAAATTAATGATGATGATGATGATGATGATGATGATGATGAT 307
QY 155 TTTATGCTCAGGCTACTCC 174
DB 308 TTTTACATTCAGGCTCTTAC 327

Search completed: June 19, 2003, 15:01:24
Job time : 123 secs

Thu Jun 19 17:30:40 2003

us-09-903-410-26.rst

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 13:31:50 ; Search time 1077 Seconds

(without alignments)
11368.420 Million cell updates/sec

Title: US-09-903-410-26

Perfect score: 756

Sequence: 1 tctgagattgaggaatttga.....tagagttacttctgatatag 756

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues 32308132

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_estin:*
3: em_estmu:*
4: em_estov:*
5: em_estro:*
6: em_estro:*
7: em_estro:*
8: em_estro:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estom:*
16: em_estom:*
17: gb_est7:*
18: em_est8:*
19: em_est9:*
20: em_est10:*
21: em_est11:*
22: em_est12:*
23: em_est13:*
24: em_est14:*
25: em_est15:*
26: em_est16:*
27: em_est17:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47.2	6.2	392	12	BF657575 OV2 23 HO
2	41.8	5.5	470	13	BM107729 C01603_CD
3	40.8	5.4	456	10	AV397710 AV397710
4	40.8	5.4	657	13	BI875128 BI875128
5	40.8	5.4	661	14	BO816618 BO816618
6	40.8	5.4	713	14	BO817181 BO817181

C	7	40.2	5.3	460	17	AZ517865	AZ517865 RPT-11-8
C	8	40	5.3	807	12	BO784079	BO784079 SEAMC004
C	9	39.8	5.3	459	17	AQ332328	AQ332328 HS_5007_A
C	10	39.8	5.3	465	17	AQ335962	AQ335962 HS_5015_B
C	11	38.6	5.1	510	17	AQ600815	AQ600815 HS_5318_A
C	12	38.6	5.1	1010	17	CNS01678	AL107180 Drosophila
C	13	38.4	5.1	361	17	AZ049266	AZ049266 GSSBR081
C	14	37.8	5.0	1244	12	AZ294693	AZ294693 4906_1455
C	15	37.6	5.0	1244	12	BF678161	BF678161 602084827
C	16	35.8	4.7	698	17	BO808806	BO808806 RPT-24-3
C	17	35.6	4.7	1101	17	CNS00001	AL065414 Drosophila
C	18	35.2	4.7	767	12	BO845295	BO845295 1024009P1
C	19	35	4.6	508	14	R30038	R30038 12643_Lambda
C	20	35	4.6	528	10	AV831633	AV831633 701558595
C	21	35	4.6	569	9	A1992658	BM479230 AGENCOURT
C	22	35	4.6	1101	13	BM479230	BO797080 EST 6018
C	23	34.8	4.6	449	14	BO797080	AZ158907 SP_0061_B
C	24	34.8	4.6	463	17	AZ158907	BO79518 EST 4086
C	25	34.8	4.6	589	14	BO79518	BO796737 EST 5675
C	26	34.8	4.6	789	14	BO796737	AA789057 ag41B12_s
C	27	34.8	4.6	816	9	AA789057	AQ531614 RPT-11-3
C	28	34.6	4.6	668	17	AQ531614	AV723426 AV723426
C	29	34.4	4.6	461	10	AV723426	AQ672914 HS_5460_A
C	30	34.2	4.5	542	17	AQ672914	AG154803 pan trogl
C	31	34.2	4.5	633	17	AG154803	BM760289 BMBA312A
C	32	34	4.5	361	17	BM760289	AI164222 A057P53U
C	33	34	4.5	408	9	AI164222	BE935559 WR2-MT012
C	34	33.8	4.5	385	12	BE935559	BO143125 fmbhc.DKO
C	35	33.8	4.5	426	14	BO143125	BO143278 fmbhc.DKO
C	36	33.8	4.5	426	14	BO143278	AZ010196 RPT-23-3
C	37	33.8	4.5	596	17	AZ010196	AL132134 Fugu tubr
C	38	33.8	4.5	619	17	FR0044642	AG077885 pan trogl
C	39	33.8	4.5	668	17	AG077885	AG103741 pan trogl
C	40	33.8	4.5	812	17	AG103741	AL076805 Drosophila
C	41	33.8	4.5	898	17	CNS0000C	AQ279281 CITBI-E1
C	42	33.6	4.4	356	17	AQ279281	AL420700 T3 end of
C	43	33.6	4.4	979	17	CNS06Y86	AM298975 180 MARC
C	44	33.4	4.4	298	10	AM298975	AM147194 dal5c07.x
C	45	33.4	4.4	451	10	AM147194	

ALIGNMENTS

RESULT 1
BF657575
LOCUS
DEFINITION
OV2.23.H05.B1_A002 Ovary 2 (OV2) Sorghum bicolor cDNA, mRNA
ACCESSION
BF657575.1 GI:11922709
VERSION
BF657575.1
KEYWORDS
EST.
SOURCE
Sorghum
ORGANISM
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE
1 (bases 1 to 392)
Cordonnier-Pratt,M.-M., Gingle,A., Marsale,C., Sudman,M. and Pratt
L.H.
An EST database from sorghum: ovaries of varying immature stages
unpublished (2000)
CONTACT
Cordonnier-Pratt MM
LABORATORY
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
TEL
Tel: 706 542 1860
FAX
Fax: 706 583 0210
EMAIL
Email: mpratt@uga.edu
Seq primer: JBN REV
High quality sequence stop: 392
POLYA-NO.

FEATURES

Location/Qualifiers

```

source
1. .392
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Ovary 2 (OV2)"
/note="Organ: Mix of ovaries of varying immature stages
from 8-week-old plants; Vector: pBluescript II from Lambda
Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made
from poly-A RNA in the cloning vector, lambda ZAP II.
Clones to be sequenced were prepared by mass excision."
BASE COUNT 79 a 94 c 104 g 115 t
ORIGIN
Query Match 6.2%; Score 47.2; DB 12; Length 392;
Best Local Similarity 56.4%; Pred. No. 0.0028;
Matches 88; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 340 GCGACGATTTATCTCGGAGGGCTCTATACCTTCGGAAGGAGTTTAATCCCGCA 399
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 220 GCCACCAATTTAAGTACGGGACGCTGAATTATGTTTACTGAAGCGCATCTCCATCTT 279
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 400 CTTCCTCGGCGAGCTGTGCAATTCGCGCATATTTCAACCCGTTGAGTATAGAAATTACTTG 459
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280 ATTCCGCGCATCATGCAATTTCCAGGACTCATGCACTGTTCACATACGCGTACTGG 339
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 460 CTGCTTGAGGAGGTATAGTTAAACAACCTTCCCGTT 495
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 340 CTGGTTGATGGAGCAGTCGTTAAACCAATTCCTATT 375
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 2
BM107729/c
LOCUS BM107729 470 bp mRNA linear EST 01-JUL-2002
DEFINITION c01e03 cDNA Century library Glycine max cDNA clone c01e03 5', mRNA
sequence.
ACCESSION BM107729
VERSION BM107729.1 GI:21637574
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE 1 (bases 1 to 470)
AUTHORS Matthews,B.F., Khan,R., MacDonald,M. and Beard,H.
TITLE Resistance mechanisms in soybean: Gene expression profile at an
early stage of soybean cyst nematode invasion
JOURNAL Unpublished (2001)
COMMENT Contact: Alkharouf, N.W.
Soybean Genomics and Improvement Laboratory (SGIL)
US Department of Agriculture (USDA), ARS, PSI
Bldg.006, Rm 118, 10300 Baltimore Ave., Beltsville, MD 20705-2350,
USA
Tel: 301 504 5750
Fax: 301 504 5728
Email: alkharouf@ars.usda.gov.
FEATURES
source
1. .470
Location/Qualifiers
/organism="Glycine max"
/cultivar="Century"
/db_xref="taxon:3847"
/clone="c01e03"
/clone_lib="cDNA Century library"
/tissue_type="cotyledons"
/dev_stage="Eight days post germination"
/note="Vector: pBluescript SK-; cDNA clones from mRNA
extracted from soybean cv. Century cotyledons 8 days post
germination."
BASE COUNT 140 a 97 c 86 g 144 t 3 others
ORIGIN
Query Match 5.5%; Score 41.8; DB 13; Length 470;
Best Local Similarity 50.3%; Pred. No. 0.14;

```

```

Matches 100; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 477 AGTTAACAACTTCCCGTTGAGCCCTTTCAGGAAACGGGTATTTCCACCGTTTGGGTTGA 536
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 395 ATTTAACAACTCTCATTTGGGTGGCAATTCAGAGGAGATTTTGTCTCCAAAGTTGAGT 336
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 537 TGTCCTTCCATAGAGCGGGAAGATATAAGAACATCTTCACATCCTCTTTGAGGAG 596
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 335 TTAATATCAAGCGGACCAATGTGTTGGAAGACATGAATCTGATTTATTGCTGAGAT 276
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 597 CTTCCTTCTTGGCGTCCGCTCAAACTCCGAAAGAGAAAGAGTTTGTGACCTCGTTAT 656
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 275 CCAATTCCTGAGGTAGCCAAATCACCAAATGTNTCAGGAAGTTCTTTGAGGCTACTGA 216
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 657 AGTTCCTGAGCTTGAGGAG 675
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215 AGTTGCTGCTCAGGTTGAG 197
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
AV397710/c
LOCUS AV397710 456 bp mRNA linear EST 23-APR-2002
DEFINITION AV397710 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii
cDNA clone CL49b11_r 5', mRNA sequence.
ACCESSION AV397710
VERSION AV397710.1 GI:6551926
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaeae; Chlamydomonas.
REFERENCE 1 (bases 1 to 456)
AUTHORS Asamizu,B., Nakamura,Y., Sato,S., Fukuzawa,H. and Tabata,S.
TITLE A large scale structural analysis of cDNAs in a unicellular green
alga, Chlamydomonas reinhardtii. I. Generation of 3433
non-redundant expressed sequence tags
JOURNAL DNA Res. 6 (6), 369-373 (1999)
MEDLINE 20152988
COMMENT Contact: Yasukazu Nakamura
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakam@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
source
1. .456
/organism="Chlamydomonas reinhardtii"
/db_xref="taxon:3055"
/clone="CL49b11_r"
/clone_lib="Chlamydomonas reinhardtii C9"
/dev_stage="photoautotrophic growth"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 101 a 142 c 136 g 77 t
ORIGIN
Query Match 5.4%; Score 40.8; DB 10; Length 456;
Best Local Similarity 59.5%; Pred. No. 0.29;
Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 460 CTGCTTACCGGAGTATAGTTACACCTTCCCGTTGAGCCCTTTCAGGAAACGGGTATT 519
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280 CTTCCTTGGCGGAGATGGCGGAGACGCTCCTCTGTTGAGGCCCTTGGGGAAGTGTCCGA 221
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 520 CCCACCGTTTGGCTTGATGTCTTCCCATAGAGCCGGAAGATATAAGAACAT 575
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 220 CTCAATGATGTCCTTGAAGCCATACCTTGTAGGCGCGCTCCAGGATCTTCTGATAT 165
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
BI875128/c
LOCUS BI875128 657 bp mRNA linear EST 11-OCT-2001
DEFINITION 963122A09.y1 C. reinhardtii CC-1690, Stress condition I, normalized

```

ACCESSION B1875128 1 GI:16073132
VERSION B1875128.1
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii;
Bukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 657)
AUTHORS Grossman, A., Chang, C.-W., Davies, J., Harris, B., Hauser, C., Lefebvre, P., McDermott, J. P., Shrago, J., Silflow, C. and Stern, D.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 963
JOURNAL Unpublished (2001)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
location/Qualifiers
1. 657
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Stress condition I, normalized, lambda Zap II"
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TRAP-N (30 min, 1hr, 4hr), TRAP-S (30 min, 1hr, 4hr), TRAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with Exsist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 143 a 198 c 191 g 125 t
ORIGIN

Query Match 5.4%; Score 40.8; DB 13; Length 657;
Best Local Similarity 59.5%; Pred. No. 0.36;
Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 460 CTGCTTGAAGAGATATAGTTAAACACCTTCCCGTTAGACCTTTTCAGAGAAAGCGTATT 519
DB 383 CTTCCTGGCGAGATGCGCGGAGCGACGCTCTCGTTAGAGCCCTTGGGAAAGTGTCCGA 324

QY 520 CCCACCGTTGCGTATGCTTCCCATAGAGCGGAAAGATATTAAGAACT 575
DB 323 CTCATGATGCTTCTGAAGCCATCTTGTAGGCGCGCTCCAGAGATCTTGAATAT 268

RESULT 5
B0816618/c 661 bp mRNA linear EST 01-AUG-2002
LOCUS 1030058H11.v1 C. reinhardtii CC-1690, Deflagellation (normalized),
DEFINITION Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION B0816618
VERSION B0816618.1 GI:22065882
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii;
Bukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 661)
AUTHORS Grossman, A., Chang, C.-W., Davies, J., Harris, B., Hauser, C., Lefebvre, P., McDermott, J. P., Shrago, J., Silflow, C. and Stern, D.

TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1030
JOURNAL Unpublished (2002)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
location/Qualifiers
1. 661
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Deflagellation (normalized), lambda Zap II"
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; Deflagellation library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells which had been re-synthesizing flagella for 15, 30 and 60 min after being deflagellated by pH shock. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with Exsist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al. (1996) Genome Research 6: 791-806."

BASE COUNT 147 a 197 c 188 g 129 t
ORIGIN

Query Match 5.4%; Score 40.8; DB 14; Length 661;
Best Local Similarity 59.5%; Pred. No. 0.36;
Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 460 CTGCTTGAAGAGATATAGTTAAACACCTTCCCGTTAGACCTTTTCAGAGAAAGCGTATT 519
DB 411 CTTCCTGGCGAGATGCGCGGAGCGACGCTCTCGTTAGAGCCCTTGGGAAAGTGTCCGA 352

QY 520 CCCACCGTTGCGTATGCTTCCCATAGAGCGGAAAGATATTAAGAACT 575
DB 351 CTCATGATGCTTCTGAAGCCATCTTGTAGGCGCGCTCCAGAGATCTTGAATAT 296

RESULT 6
B0817181/c 713 bp mRNA linear EST 01-AUG-2002
LOCUS 1030062C12.x1 C. reinhardtii CC-1690, Deflagellation (normalized),
DEFINITION Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION B0817181
VERSION B0817181.1 GI:22066838
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii;
Bukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 713)
AUTHORS Grossman, A., Chang, C.-W., Davies, J., Harris, B., Hauser, C., Lefebvre, P., McDermott, J. P., Shrago, J., Silflow, C. and Stern, D.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1030
JOURNAL Unpublished (2002)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
location/Qualifiers

source

1. .713
 /organism="Chlamydomonas reinhardtii"
 /strain="CC-1690 wild type mt+ 21gr"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, Deflagellation
 (normalized), Lambda Zap II"
 /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
 XhoI; Deflagellation library, constructed by John Davies
 and Jeffrey McDermott, combines cDNAs from CC-1690 cells
 which had been re-synthesizing flagella for 15, 30 and 60
 min after being deflagellated by pH shock. PolyA mRNA was
 purified from each sample, pooled and cDNA synthesized.
 The cDNA was directionally cloned into lambda Zap II
 (Stratagene) in the EcoRI (5') and XhoI (3') sites.
 pBluescript II SK- plasmids were excised from the lambda
 Zap clones by superinfection with ExAssist (Stratagene)
 phage. The library was normalized using method 4 described
 in Bonaldo et al., (1996) Genome Research 6: 791-806."
 BASE COUNT 153 a 216 c 204 g 140 t
 ORIGIN

Query Match 5.4%; Score 40.8; DB 14; Length 713;
 Best Local Similarity 59.5%; Pred. No. 0.38;
 Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 460 CTCGTTGACGAGTATAGTTAAACACCTTCCCGTTGAGCCCTTTCAGGAAAGCGTATT 519
 |||||
 Db 439 CTTCTTGGCGAGATGGCGGACGACGTCCTCGTTGAGGCCCTTGGGGAAGTGTCCGA 380
 |||||
 QY 520 CCCACCGTTGCGTTGATGTCCTTCCCATAGACCGGAAAGGATATAAAGACAT 575
 |||||
 Db 379 CTCATGATGTCCTTGAAGCCATCTTGAGGCCGCGTCCAGGATCTCTGAATAT 324
 |||||

RESULT 7

AZ517865/c
 LOCUS 460 bp DNA linear GSS 16-OCT-2000
 DEFINITION RPCI-11-89H18.TVB RPCI-11 Homo sapiens genomic clone RPCI-11-89H18,
 DNA sequence.

ACCESSION AZ517865
 VERSION AZ517865.1 GI:10827791
 KEYWORDS GSS.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 460)
 Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
 J.C.

REFERENCE

AUTHORS
 TITLE BAC end sequences of library RPCI-11
 JOURNAL Unpublished (1997)
 COMMENT Other GSSs: RPCI11-89H18.TJ RPCI11-89H18.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute of Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

Email: szhao@tigr.org
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
 Research Genet cs (info@resgen.com). BAC end search page:
 http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
 This BAC end was generated during the R&D process and may have
 higher chance of clone tracking errors.
 Seq primer: T7
 Class: BAC ends.

FEATURES

source

Location/Qualifiers
 1. .460
 /organism="Homo sapiens"
 /db_xref="GDB:753397"

BASE COUNT 146 a 79 c 95 g 140 t
 ORIGIN

Query Match 5.3%; Score 40.2; DB 17; Length 460;
 Best Local Similarity 56.4%; Pred. No. 0.44;
 Matches 75; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 555 GGAAGAAGGATATAAGAACATCTTCACATCTCTTTGAGGAGCTTCTTCTGCGTCCG 614
 |||||
 Db 215 GCAGAGGATACAGATAGACATCACTATTAGTTGTACATATCTTTTGGTGT 156
 |||||
 QY 615 CTCAAACTCCGAAAAGAGAAAGAGTTTGTGACCTCGTTATAGTTCTCTGAGCTTGAAGA 674
 |||||
 Db 155 TTCAACTTGTGAAATGTCTATTCTGCTGTTTTCAGCTTGCATGAACTTGATGAGTTGAAGA 96
 |||||
 QY 675 GTTCACACCCCTT 687
 |||||
 Db 95 GGTCAACCAAGTT 83
 |||||

RESULT 8

BG784079
 LOCUS 807 bp mRNA linear EST 20-MAY-2001
 DEFINITION SEAMC004036 Sea urchin primary mesenchyme cell cDNA library
 Strongylocentrotus purpuratus cDNA clone PC_0020_A1_B12_MR 5', mRNA
 sequence.

ACCESSION BG784079.1 GI:14155092
 VERSION BG784079
 KEYWORDS EST.
 ORGANISM Strongylocentrotus purpuratus.

Strongylocentrotus purpuratus
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 Echinoidea; Euechinoidea; Echinacea; Echinoida;
 Strongylocentrotidae; Strongylocentrotus.
 1 (bases 1 to 807)
 Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and
 Ettensohn,C.A.

REFERENCE

AUTHORS
 TITLE A large scale analysis of mRNAs expressed by primary mesenchyme
 cells of the sea urchin embryo
 JOURNAL Development 128 (13), 2615-2627 (2001)
 MEDLINE 21384984
 COMMENT Contact: Ettensohn CA
 Dept. Biol. Sci.
 Carnegie Mellon University
 4400 Fifth Avenue, Pittsburgh, PA 15213, USA
 Tel: +1 412 268 5849
 Email: ettensohn@andrew.cmu.edu.

FEATURES

source

Location/Qualifiers
 1. .807
 /organism="Strongylocentrotus purpuratus"
 /db_xref="taxon:7668"
 /clone="PC_0020_A1_B12_MR"
 /clone_lib="Sea urchin primary mesenchyme cell cDNA
 library"
 /tissue type="embryo"
 /cell type="primary mesenchyme cells"
 /lab host="E.coli"
 /note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; oligo
 at priming from poly A+ RNA, directionally cloned"
 BASE COUNT 197 a 209 c 209 g 191 t 1 others

Query Match

Best Local Similarity 5.3%; Score 40; DB 12; Length 807;
 Matches 67; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 51 TGCMAAGGCGATACCCGACATAGTGTGTTGAAAGCTATATAACGAGCTCGTATTAAGGGT 110
 |||||
 DB 44 TGCMAAGGCGATATATCCAGCTCGGACATCATCAAGAGAGTACGATCCCAT 103
 |||||
 QY 111 GAGGCGTTTAAAGCGGGGTAGCGCGCGGCGCATCGTTTGGCTTTTATAGCC 162
 |||||
 DB 104 CGACATGTCGTGGCGGGGTACGACATCGGGGCCCTTCATCTCGCCCTTACGCC 155
 |||||

RESULT 9
 A0332328 459 bp DNA linear GSS 06-MAR-1999
 LOCUS HS_5007_A2_H07_T7 RPC111 Human Male BAC library Homo sapiens
 DEFINITION genomic clone Plate=583 Col=14 Row=O, DNA sequence.
 A0332328
 VERSION A0332328.1 GI:4129955
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 459)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 Title Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
 99380589
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 583 row: O column: 14
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 459.
 Location/Qualifiers
 1..459
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=583 Col=14 Row=O"
 /clone_1ib="RPC111 Human Male BAC library"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBAC3.6; RPC111 Human Male BAC library"

BASE COUNT 147 a 78 c 159 t 1 others

ORIGIN

Query Match 5.3%; Score 39.8; DB 17; Length 459;
 Best Local Similarity 49.8%; Pred. No. 0.58;
 Matches 101; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 199 AAGAGGTTAACTGCTGAAGCTGTTAAGTTCAGCCACTCTGAAGGATGATAGG 258
 |||||
 DB 272 AAAAGGTCAGAGCTCAATTCATTTTCAGTTTAACTATATTTTAAAAAGATGA 213
 |||||
 QY 259 TGGAGAGGCTTAAAGTCTTGAAGAGTTCCTCCCTTACAGGAGATAGAAAATT 318
 |||||
 DB 212 TTTGGGTTAAAAATAAGATGCGTTAGTTTATCTCCCTTACAGTAAAGCAAACT 153
 |||||
 QY 319 GAGATACCGAGCTATATATGCGGAGGATTTATCTCGGAGAGGCTCTATACCTCG 378
 |||||
 DB 152 GAGCTGAAAACACATAGAGCTACTCATTTCTATTTCACAAAGAACTGTTACTGTC 93
 |||||
 QY 379 GAAGGAGTTTAAATCCCGCACT 401
 |||||
 DB 92 GTATGCAAAATATATCTCTCAGT 70
 |||||

RESULT 10
 A0335962 465 bp DNA linear GSS 06-MAR-1999
 LOCUS HS_5015_B1_E07_S66E RPC111 Human Male BAC library Homo sapiens
 DEFINITION genomic clone Plate=591 Col=13 Row=J, DNA sequence.
 A0335962
 VERSION A0335962.1 GI:4142941
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 465)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 Title Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
 99380589
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 591 row: J column: 13
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 465.
 Location/Qualifiers
 1..465
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=591 Col=13 Row=J"
 /clone_1ib="RPC111 Human Male BAC library"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBAC3.6; RPC111 Human Male BAC library"

BASE COUNT 149 a 76 c 162 t 2 others

ORIGIN

Query Match 5.3%; Score 39.8; DB 17; Length 465;
 Best Local Similarity 49.8%; Pred. No. 0.59;
 Matches 101; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 199 AAGAGGTTAACTGCTGAAGCTGTTAAGTTCAGCCACTCTGAAGGATGATAGG 258
 |||||
 DB 246 AAAAGGTCAGAGCTCAATTCATTTTCAGTTTAACTATATTTTAAAAAGATGA 187
 |||||
 QY 259 TGGAGAGGCTTAAAGTCTTGAAGAGTTCCTCCCTTACAGGAGATAGAAAATT 318
 |||||
 DB 186 TTTGGGTTAAAAATAAGATGCGTTAGTTTATCTCCCTTACAGTAAAGTACGAAAATT 127
 |||||
 QY 319 GAGATACCGAGCTATATATGCGGAGGATTTATCTCGGAGAGGCTCTATACCTCG 378
 |||||
 DB 126 GAGCTGAAAACACATAGAGCTACTCATTTCTATTTCACAAAGAACTGTTACTGTC 67
 |||||
 QY 379 GAAGGAGTTTAAATCCCGCACT 401
 |||||
 DB 66 GTCTGCAAAATATATCTCTCAGT 44
 |||||

RESULT 11
 A0600815 510 bp DNA linear GSS 10-JUN-1999
 LOCUS HS_5318_A2_C09_T7A RPC1-11 Human Male BAC library Homo sapiens
 DEFINITION genomic clone Plate=894 Col=18 Row=E, DNA sequence.

VERSION A2049266.1 GI:7273181
 KEYWORDS GSS.
 SOURCE Brucella melitensis biovar Abortus.
 ORGANISM Brucella melitensis biovar Abortus
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Brucellaceae; Brucella.
 REFERENCE 1 (bases 1 to 361)
 Sanchez,D.O., Zandomeni,R.O., Cravero,S., Verdun,R.E., Pierrou,B.,
 Faccio,P., Diaz,G., Lanzavecchia,S., Agüero,F., Frasch,A.C.C.,
 Anderson,S.G.E., Rosetti,O.L., Grau,O and Ugalde,R.A.
 Gene discovery through genomic sequencing of Brucella abortus
 Infect. Immun. 69 (2), 865-868 (2001)
 TITLE 21101034
 JOURNAL Contact: Zandomeni, R.
 MEDLINE Centro de Investigación en Ciencias Agropecuarias (CICA)
 Instituto Nacional de Tecnología Agropecuaria (INTA) C.C. 25 (1712)
 Castellar, Buenos Aires, Argentina
 Tel: 5411-4621-3316/1683
 Fax: 5411-4481-1316
 Email: zandomeni@inta.gov.ar
 Class: shotgun.
 FEATURES
 source Location/Qualifiers
 1..361
 /organism="Brucella melitensis biovar Abortus"
 /strain="S-2308"
 /db_xref="taxon:235"
 /clone="B23"
 /note="lib="Sheared genomic library,"
 /note="Vector: pBluescript SK(-) (STRATAGENE); Genomic DNA
 was mechanically sheared, blunt ended, and
 size-fractionated by agarose gel electrophoresis.
 Fragments between 1.5-3 kb were recovered and ligated to
 the EcoRV site of the pBluescript SK (-) vector."
 BASE COUNT 88 a 112 c 97 g 63 t 1 others
 ORIGIN
 Query Match 5.1%; Score 38.4; DB 17; Length 361;
 Best Local Similarity 47.8%; Pred. No. 1.4;
 Matches 111; Conservative 0; Mismatches 121; Indels 0; Gaps 0;
 QY 308 TAGAAAACTTGAGATACCACTATATATGCGGACGATTATATCTGGGAGAGCGCTC 367
 DB 245 TGAAGACCTGGAATATCCCATGACATTAACGCGGTGATTTTCATAGCGCGCGGAGAA 186
 QY 368 TATACCTCTCGAGAGGAGTTTATCCCGACCTTCGCGACGCTGTGCAATCCCGGGA 427
 DB 185 TTCAATAGCGGATGGATTTACGCTCGGCGCAAGCTGCTCGCCCATTCACCCG 126
 QY 428 TATTGAACCGCTTGAGATATAGAAATTAATCTGCTGTGACGAGGTATATTAACACCC 487
 DB 125 TTTTGGCGCGGTGGCGCGCATGGCGCGCATCTGGTGCATGGCGGGCTGTTCAATCCG 66
 QY 488 TTCCCGTTGAGCCCTTTCAGAGAAAGGGATATTCACACCGTTTGGCTTATGT 539
 DB 65 TGCGGTGACCTCTTTTCACACAGCGGATATCGTATCGGCATCGATGT 14
 RESULT 14
 A2924693/c 462 bp DNA linear GSS 01-APR-2001
 LOCUS 4906.id5b14.s1 Saccharomyces mikatae IFO 1815 Saccharomyces
 DEFINITION mikatae genomic clone 4906.id5b14.s1, DNA sequence.
 ACCESSION A2924693
 VERSION A2924693.1 GI:13495592
 KEYWORDS GSS.
 SOURCE Saccharomyces mikatae.
 ORGANISM Saccharomyces mikatae
 Bacteria; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 REFERENCE 1 (bases 1 to 462)
 Cliften,P.F., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Gish
 W.R., Waterston,R.H. and Johnston,M.
 Surveying Saccharomyces genomes to identify functional elements by

JOURNAL comparative DNA sequence analysis
 COMMENT Unpublished (2001)
 CONTACT: Johnston M
 Department of Genetics
 Washington University Medical School
 Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
 Tel: 314 362 2735
 Fax: 314 362 7855
 Email: wj@genetics.wustl.edu
 Classes: random plasmid subclone.
 TITLE
 JOURNAL
 MEDLINE
 FEATURES
 source Location/Qualifiers
 1..462
 /organism="Saccharomyces mikatae"
 /strain="IFO 1815"
 /db_xref="taxon:114525"
 /clone="4906.id5b14.s1"
 /clone_lib="Saccharomyces mikatae IFO 1815"
 /note="Random genomic sequence"
 BASE COUNT 142 a 107 c 72 g 141 t
 ORIGIN
 Query Match 5.0%; Score 37.8; DB 17; Length 462;
 Best Local Similarity 54.7%; Pred. No. 2.4;
 Matches 75; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
 QY 10 AGGAATTGGAAGATTAACCTGCTCTTCGCGAGAGCTGCAAGGCGATAGCCAC 69
 DB 266 AGGATATCGCTGCGGACCGCTTTAGTTTGTGAGGTGTACTTTTGCTTTTCAC 207
 QY 70 ATAGGTGTTTGAAGATTAAGAGAGCTCGGTATTAAGGTGAGCGCTTTAAGCGGGTGG 129
 DB 206 ATGTGTGTTCTTGGATGTTATGGAATTAATTAATTAATTAATTAATTAATTAATTAAT 147
 QY 130 AGCGCCGGGGCAATGCT 146
 DB 146 AGTGCGGTGCAATGT 130
 RESULT 15
 BF678161 1244 bp mRNA linear EST 21-DEC-2000
 LOCUS 602084827P1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4248967 5',
 DEFINITION mRNA sequence.
 ACCESSION BF678161
 VERSION BF678161.1 GI:11952056
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1244)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cga@bbs-rtmail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LNC1068 row: f column: 08
 High quality sequence start: 6
 High quality sequence stop: 88.
 TITLE
 JOURNAL
 MEDLINE
 FEATURES
 source Location/Qualifiers
 1..1244
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4248967"
 /clone_lib="NIH_MGC_83"
 /lab_host="DH10B (TI phage-resistant)"

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 13:05:45 ; Search time 1449 Seconds

(without alignments)
15184.073 Million cell updates/sec

Title: US-09-903-410-26

Perfect score: 756
Sequence: 1 ttgagatcgaggaaatttga.....tagagttactttcgaatag 756

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	403.6	53.4	15820	1	AE000737	AE000737 Aquifex a
2	63.8	8.4	11357	1	AE012939	AE012939 Chlorobiu
3	57.2	7.6	11039	1	AE004756	AE004756 Pseudomon
4	55.4	7.3	11618	1	AE013136	AE013136 Thermocana
5	55.2	7.3	2797	1	ECJ224437	AE224437 Erwina c
6	55	7.3	329100	1	SME591787	AL591787 Sinorhizo
7	53.6	7.1	903	6	A96528	A96528 Sequence 56
8	53.6	7.1	340806	1	NMA1Z2491	AL162752 Neisseria
9	52	6.9	903	6	A96526	A96526 Sequence 55
10	52	6.9	9831	1	AE002562	AE002562 Neisseria
11	52	6.9	172325	6	AX044035	AX044035 Sequence
12	52	6.9	300950	1	AP001516	AP001516 Bacillus
13	49.8	6.6	10410	1	AE005340	AE005340 Escherich
14	49.8	6.6	291136	1	AP002556	AP002556 Escherich
15	48.2	6.4	3653	1	ECOTGSA	M64675 Escherichia
16	48.2	6.4	10648	1	AE000221	AE000221 Escherich
17	48.2	6.4	13860	1	D90758	D90758 Escherichia
18	48.2	6.4	16397	1	D90759	D90759 Escherichia
19	48.2	6.4	23203	1	D90852	D90852 E.coli geno
20	48	6.3	4406	1	EC410309	MJ410309 Erwina c
21	47.4	6.3	903	6	A96530	A96530 Sequence 56
22	47.2	6.2	2266	1	SFTYRT	X66849 S.flexneri
23	46.6	6.2	22388	1	AE008778	AE008778 Salmonell
24	46.6	6.2	254050	1	AL627269	AL627269 Salmonell
25	46.2	6.1	40138	6	AR118084	AR118084 Sequence
26	46.2	6.1	41587	1	AB025342	AB025342 Moricella
27	46	6.1	998	9	HS435659	MJ325659 Homo sapi
28	45.8	6.1	467	6	AX438333	AX438333 Sequence
29	43.6	5.8	218470	1	BSUB0013	D84432 Bacillus su
30	43.6	5.8	282700	1	BACJH642	AP002996 Mesorhizo
31	43.6	5.8	349619	1	AP002996	AE008096 Agrobacte
32	42.8	5.7	10741	1	AE008096	AE009131 Agrobacte
33	42.8	5.7	11010	1	AE009131	AL590445 chromosom
34	42.8	5.7	211018	3	CNS07BGC	A96524 Sequence 55
35	42.6	5.6	447	6	A96524	AE003860 Xylella f
36	42.6	5.6	10593	1	AE003860	AE008531 Brucella
37	40.8	5.4	10029	1	AE009531	AE009531 Sequence 14
38	40.4	5.3	7218	6	166494	AC105161 Mus muscu
39	40.2	5.3	253946	2	AC105161	AC023142 Homo sapi
40	39.8	5.3	159505	2	AC023142	AF178030 Homo sapi
41	39.8	5.3	310001	9	AF178030	AF404759 Bacteroid
42	39.6	5.2	5137	1	AF404759	AX067466 Sequence
43	39.6	5.2	269223	6	AX067466	AF026065 Ralstonia
44	39	5.2	9511	1	AF026065	B22111 Aromatic co
45	39	5.2	12017	6	E22111	

ALIGNMENTS

RESULT 1
AE000737
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

AE000737 15820 bp DNA linear BCT 25-MAR-1998
Aquifex aeolicus section 69 of 109 of the complete genome.
AE000737.1 GI:2983782
Aquifex aeolicus.
Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
Aquifex aeolicus
1 (bases 1 to 15820)
Deckert,G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L.,
Graham,D.E., Overbeek,R., Sneed,M.A., Keller,M., Ajay,M.,
Huber,R., Feldman,R.A., Short,J.M., Olsen,G.J. and Swanson,R.V.
The complete genome of the hyperthermophilic bacterium Aquifex

aeolicus
 JOURNAL Nature 392 (6674), 353-358 (1998)
 MEDLINE 98196666
 PUBMED 9537320
 REFERENCE 2 (bases 1 to 15820)
 AUTHORS Deckert, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Overbeek, R., Snead, M. A., Keller, M., Aujay, M., Huber, R., Feldman, R. A., Short, J. M., Olson, G. J. and Swanson, R. V.
 TITLE Direct Submission
 JOURNAL Submitted (25-JUL-1997) Diversa Corporation, Genomics, San Diego, CA 92121
 COMMENT Putative indicates no similarity to known proteins
 FEATURES Hypothesis indicates similarity to a protein of unknown function.
 Location/Qualifiers
 1..15820
 /organism="Aquifex aeolicus"
 /strain="VF5"
 /db_xref="taxon:63363"
 complement (12..1298)
 /gene="gpsi"
 /note="aq_1367"
 complement (12..1298)
 /gene="gpsi"
 /codon_start=1
 /transl_table=11
 /product="glucose-1-phosphate thymidyltransferase"
 /protein_id="AAC07344.1"
 /db_xref="GI:2983788"
 /translations="MVEAVILAGGEGNRLKPTVEVPKALLKVGARELLYRTIKQLQ
 DVGKFNVIYVNNKFGKVKAFLEKHNFEAEVIPNHPKENGYSLYLAKRIKGEFA
 VNSDHIKFALEKAVEGKGLIVDRGLYINKNEATKVKCEGRKIKYIGKLEKYDG
 FDTGFFVLDESIFEVAEALKEQKLTMSLAKRAOIPTCEVSGVPMVDVTPDEVEK
 AKYLVTALIKGVDGFIENLARKVSTRISPLVDKFTPNQLTVLTLLGMPALVA
 YESPALGILLQINSLMDGLDGEVARAQRNTKFGAWLSDVDRYVDFLALSAWLH
 KPSWFMFWALFAGSVWVSYSYTERYKAGYCEDAYAVIKELRYLLGKRDRIFFIMI
 FTILGWIKALFVLAITLNRVLTITLYLVNKKGVN"
 /gene="tagD2"
 /note="aq_1368"
 complement (1313..1819)
 /gene="tagD2"
 /note="aq_1369"
 complement (1313..1819)
 /gene="tagD2"
 /codon_start=1
 /transl_table=11
 /product="glycerol-3-phosphate cytidyltransferase"
 /protein_id="AAC07343.1"
 /db_xref="GI:2983785"
 /translations="MNNKGVVITYGTDFLHGHNLKRAKALGDFLIVGVSTDEF
 NAIKGGKSVYPYEHRAEIVRSIKYVDLVIPERNWEQIEDIKYNDVPMGDDWKGK
 FDLKEYCEVYVLPRTREGISTTELKEALVKLSNILEEDILRLVDIWRGRICIEILK
 NKNLKNGE"
 1794..2576
 /gene="aq_1369"
 1794..2576
 /gene="aq_1369"
 /codon_start=1
 /transl_table=11
 /product="putative protein"
 /protein_id="AAC07356.1"
 /db_xref="GI:2983798"
 /translations="MTTFPFLFIVLPYTLMPFVKIFLKLGRDMEIQTLVRSDDKL
 TPNREFLKNPEYKYLKINWSELQGFQKLGNEYLVIYIRPEKPEKELGIF
 QSVREAMGAPLSTALEHGWEEVPKNYVIYHADPVEGNGKLAIAIKTEGISTYDQKL
 ENMKMWRVPRVYSSDVLTKIDIPDVQSKAVIAREIAKETGSAPELEELGKI
 YGVDTSGLSGLELIEKLLQNPVKLPGGKVENLKYWYFLEA"
 2573..3088
 /gene="ripA1"
 /note="aq_1370"
 2573..3088
 /gene="ripA1"
 /codon_start=1
 /transl_table=11
 /product="rare lipoprotein A"
 /protein_id="AAC07342.1"

/db_xref="GI:2983784"
 /translation="MIALIFPLILSFVSVKELFLEDPLKEPSEPKLFLKEBRYKRI
 RIFKVKDCRYEVGVYASWYGPFKYRTANGERNFMFKLTAASTFSLGTYLVVYNLE
 NGKFTVVRINDRGYPVDGRIIDLSMAAADIEMLKKGKGVAKVLVPLKCLAPSTQVKY
 BEVVDIMKTY"
 3099..4472
 /gene="argH"
 /note="aq_1372"
 3099..4472
 /gene="argH"
 /codon_start=1
 /transl_table=11
 /product="argininosuccinate lyase"
 /protein_id="AAC07341.1"
 /db_xref="GI:2983783"
 /translations="MEKFWSGRPFKEETDKFVEDFTESVDFDKELAFEDIEODIAHVKT
 LOKAGILTEERARELIQRLKIKEIKEGKFWKKELEDVHMNTEAEILNRLGVDGRK
 LHTARNDOVATDEKLYLKEIKELVQLLKLKTKTVELAETTVDFVMSYTHLQRA
 QPIRVAYFLAYREILLKDSERLMDTRRDEDELPLGSGAVAGVDFFDRFTYALLGF
 NRVTRNSMTAERDFFIEFNSCALIAQHLRLAELDIINWTEEFNFVELPDKLCTG
 SSIMPOKNPDVLEIRKGTIRYGNLIALLITMKALPMAYNEDMDKDFLPTLKN
 LKNNIKGMLVLSLRVKEONMRKASGNFLITDIANYLVKGVPPRTAHIAIGSIVA
 YLEKGGKLEEMTLEEFKQFSEKPEEDVDFILSPERAADRKRVVGGTAKKEILLRIEV
 AKAEGR"
 4548..5327
 /gene="nuoH2"
 /note="aq_1373"
 4548..5327
 /gene="nuoH2"
 /codon_start=1
 /transl_table=11
 /product="NADH dehydrogenase I chain H"
 /protein_id="AAC07347.1"
 /db_xref="GI:2983789"
 /translations="MHSFWINLLPAVIKIVILLIAGIGAFLTWVERKVAAHIOREP
 GPMVWGHLLOPLANGLKLLTKEDLPRYGDPLFHLHLAILALVPATLIVSVVPFGP
 EFEPYGVKYPILSDVNVGLLLPALGSMAYVALGASNSKIPILASMRKAGVLV
 SIEVITFAVGPIMLAGTISTYEIVQKIEQNLWYIWOQPIAFVVTVMFAALAEGRV
 PFDVQEAELVTGFTVEYGGMKFGLPFLVEWYIEVLPCLL"
 5294..5596
 /gene="nuoH3"
 /note="aq_1374"
 5294..5596
 /gene="nuoH3"
 /codon_start=1
 /transl_table=11
 /product="NADH dehydrogenase I chain H"
 /protein_id="AAC07348.1"
 /db_xref="GI:2983790"
 /translations="MWRSLTSLAIVLVFFGWSPINIPFVGFDPLFLGLPLSPV
 WFLIKTALFLFILMLHWTLPYRDIQITQNAWKIMLPLTLFNLVLAFLAVWR"
 5603..6229
 /gene="nuoI2"
 /note="aq_1375"
 5603..6229
 /gene="nuoI2"
 /codon_start=1
 /transl_table=11
 /product="NADH dehydrogenase I chain I"
 /protein_id="AAC07349.1"
 /db_xref="GI:2983791"
 /translations="MIKKVAKPLSWLERIFFIDFKGLRITLKNALRKTITTHYPYB
 KIPPKRFGTFAHKVVDGTEPQPAFOEWNRYNIIIVYEGKSCVCLCRKACVPQ
 LFIIEGKKLPNGKRVSVFNMMLLCTYCGFCVADPCDCLYQTDIHNASVTKDAV
 LTLLEIQIGRDWRREPRDPRIWIDDEORMKLGNNVNLKPKBEV"
 6226..6762
 /gene="nuoJ2"
 /note="aq_1377"
 6226..6762
 /gene="nuoJ2"
 /codon_start=1
 /transl_table=11
 /product="NADH dehydrogenase I chain J"


```

/codon_start=1
/transl_table=11
/product="type I restriction system endonuclease"
/protein_id="AAW73097.1"
/db_xref="GI:21647897"
/translation="MWSRMTDKYTKIAESNNFVILDRYMPDMKVAEGYQSEADLERELI
DLRRGQYDFPAIKTEPEAMLANVRVQLANDVQFSDGEWARFVETVLDPSGIVE
KTRKHIDDYIHFDVDDGRIGIQHILWKKTKLVNRKQVIRVQVFTTTPALCATPPREGNF
S3GEQLMDPPGPGVGVGKGVYDVTILVNGLPVQVELKRGVVAIRAFANQVHR
YKSEFNSAHLFKYLQLFVINGDTRVFANTTRDKNSDPFTMHAKADNTIRDL
KOPAAFPQKHLLSVLHYSVDFVSNLTLLVMRPYQIAATERILWKIKSSHQAKTWS
PEGGGIWHITSGKLTSLFKARLGTDFIDKVFVVDKDLDTQMKYQRPSPD
SVNGSDTAGLRNLEKDNRIIVTIOKLNMKSEBPLPIYHKQVVFIDECHRSQ
F3EAKLKKTKRYLQFGTGTPTIPPENALGAETASVFGRELSYVITDAIRDEKY
LKFQVNDVRFRRKLAQFTEDEKLSAENRQALLHPERIREITQYILTHYOKTHR
LOPAGFNALFVSVSVAEAKLYEAFKQKDSAKPLKATIFSYAENEAQDVGDI
ADGFEVSLNSAKEFLNAAIADYNALFKTNFSDVSQGFNYRDLAKVGTDDSG
KLPADEQVLLIVGMFTGDAFTLNTLFDKLNRLTHGLLQASRNRIFDANKTF
GNIVTRDLEQATIDAITLFGDKNTRNVVLEKSYREYMEGYTDALTGOARRGFVEVQ
ELQARPPDPALEKEADKAFVRLFCGYLRAENVLQNFDEFAALKALQSVNTGPPAAV
EAFKQHYLSDSDLAALQAIKLPPERTMODYSTYNDIRMLRREQAQGVKESKSIDW
DVVPVDLLKSOENLDYILELIPERNKETS KAELVEEVRVIRASLGHAKESLV
VDFINQTDLEQADKASVLEAFETPARAELQEAQELIEAKLNAEAARYIATSLR
ERASDTGLNAVLPMSPLNFQYLTQKQSVFQKIAAFVEKFGVGGV"
/complement(3322..3903)
/genes="Ct1879"
/complement(3322..3903)
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAW73098.1"
/db_xref="GI:21647898"
/translation="MKETQYKSLVPNPAIRDRAKARKAGILHEALLWFLKSNKLM
GLDEPQKLIQNYIVDFYCAERSVIEVDGSHDSKQIEDREDRDLNGLGLTIVRL
ADVLNRLGQVVEFLKDHPLALGTTPPEGKNTALGTTPPEGKNTALGTTPPEGKNT
ALTGTPPEGKNTALGTTPPEGKNTTLTEARE"
/complement(3896..5320)
/genes="hads-2"
/notes="hads-2"
/complement(3896..5320)
/genes="Ct1880"
/notes="hads-2"
/notes="identified by match to PFAM protein family HMM
PF01420"
/codon_start=1
/transl_table=11
/product="type I restriction system specificity protein"
/protein_id="AAW73099.1"
/db_xref="GI:21647899"
/translation="MSGREFLQKLIDGERVEWKALGETIQLEKQQLNKLSSSGRY
PAYNGMSYSGFTDSYNSYENKTIISQGSAGSFYVFTTKFYANAHCVVLPDTEVV
DNRYIYHFLKLAERLTSQHGAGIPALRASEITSIKIPIPCDNPKKSLAIOAIRV
ILDFTALTELTARLKOVAYVRDLRLTFTTPPYGHPKSGGELSLRHPSEGG
ELFTPYGHSVEERELNSPSLKGWQAPQDGVPEWKTGLEVGHFIRSGIOKSPKAS
GVCYIHGYHGYHTWTETKSPIDPEPANRLKAKPGDLVIATISDEDDAVAKAVA
WGTEDVASTDIYIRHPANPKYMSIFFQTMDFQEQKPYITGKVRIRSGDNLAKI
LIPILPLASQERIVALLDQFDALNTSLTEGLPRELRLQKQYAYVRDLLFSPKASF
GVEPGRDQPSFGVPEGRGLNA"
/complement(5487..7043)
/genes="hadsM-2"
/notes="hadsM-2"
/complement(5487..7043)
/genes="Ct1881"
/notes="hadsM-2"
/notes="identified by match to PFAM protein family HMM
PF02384"
/codon_start=1
/transl_table=11
/product="type I restriction system adenine methylase"
/protein_id="AAW73100.1"
/db_xref="GI:21647900"
/translation="MTSIQORAEQLQRRIWQIANDVRGVTDGWDFFQYVGLAFYRFPIS
ENFAHMEAGDDGIRYAEIPLDSVITPELNDDAIKTKGYFIYPSQLFANVAVANTNDS
LNTDLAAIPTAIESSANGPYSEQDIKGLFADFTTNSRLGNTVKDKNQLAAVLKGV
ELDGPDDAHIDLFGDAYEFLISNYAANAGKSGGEFTTPOHVSRLIARLALHGOXSV
NKIYDPACSGSLLLOAKKPPEDRLIEDGFRGQESNHTTNYLARMWFLHNIYDKFN
IQCNITLLEPHFADEKFPDAIVSNPPYSVKWIGSDPTLINDERFAPAGVLAPKSKAD
FATVHALHYSKAGRAAIVFPFIFRGAAEKIRQYILVDNNYVETVIALAPNLFPG
TTIATVNLVLSKHPDPTTQFIDASALFKETNNVLLDEHIEQIMAVFASKEEVPHV
AQGVPLERTAAANNYLSVSSVEARDTREVDVIAQLNAELKTTVARIDELRKQIDAIV
AETGSEDEDA"
/complement(7106..8236)
/genes="Ct1882"
/complement(7106..8236)
/genes="Ct1882"
/notes="identified by match to PFAM protein family HMM
PF00534"
/codon_start=1
/transl_table=11
/product="glycosyl transferase"
/protein_id="AAW73101.1"
/db_xref="GI:21647901"
/translation="MKVALYAGTYVKDQAVRSIYQLVSSMIKNGHQVWVWTPDFTT
EAKSVPNLTPSPVPIPLPYDYKLGFPVNAVTERQLOEFAPDI VHI STPDIVGRKFLHY
AKKGLPVGSVAVHTDPESVLSYVRLGFAEPVAVRFLKPVNACDVTLAPNESVRELT
GKGIERVELMSRGIDKELDPDSRRSAKLRRAMDARTVIIYAGRVFLYKIDBVMSL
YQPADEGLDGRVRFVMI GSGPEEAQRRAMPVAVFTGYLTGTTLPEAVASGDLFLFP
STTFAFCNVTLALATGLPAVSDVCGQCELVSGGFGVAKAGDVGDVDFYACCTKMQ
DGBLFRSMRERGLAFADKSWAAVNGKLDIYRLIELIAAKARR"
/complement(8408..8875)
/genes="Ct1883"
/complement(8408..8875)
/genes="Ct1883"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAW73102.1"
/db_xref="GI:21647902"
/translation="MPIRRLRPLDQSHQVYFVISHSPAYTAODIAAAHVSGNELVK
TVNVSIDGKNAMALLHPRHLDLRLCELGSRDVTLAEIAPSGFGECEIGAMPFF
GNLYGKMYVADELDSEMDIVFNAGTHRELLRLSWFDYKRLVNPVVMGRIASIR"
/8875..9003
/genes="Ct1884"
/8875..9003
/genes="Ct1884"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAW73103.1"
/db_xref="GI:21647903"
/translation="MRPPWPCVPEPLMFTBHKFOSVEPARNEKPARHVMVTGRRNP"
/complement(9084..9866)
/genes="Ct1885"
/complement(9084..9866)
/genes="Ct1885"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAW73104.1"
/db_xref="GI:21647904"
/translation="MSTTENVSBRTGLAFGGVVLGAHIGVLKAMEETCPRAECVSG
TSIGSTIANYAFKSWREIREVALELDWSDLSGLTSLGSLTSIRKPKIVRAQLGS
RRIEDAPLPAIVATDICTGNEVAREVALELDWSDLSGLTSLGSLTSIRKPKIVRAQLGS
LTENVPVSLPKMGASRIACIDLFGRHSFRPRPEHLSDLLINAFYSAMRAISQIQISKA
DLTVAPDLRSLVDMDSAVPEILDITGTYREALPLLESWRDAHR"
/complement(9871..11253)
/genes="alda"
/notes="alda"
/complement(9871..11253)
/genes="Ct1886"
/complement(9871..11253)
/genes="alda"
/notes="identified by match to PFAM protein family HMM
PF00171"
/codon_start=1

```



```
/transl_table=11
/product="aldehyde dehydrogenase"
/protein_id="AA004756.1"
/db_xref="GI:21647905"
/translation="MTYDASTAGLRKYDPSGDTPEMRRAQLGLAPLRERHEI
AAVHADLRKPVATMTLPTALRSEIRIVLKLRLMRPKVGVPLIYQPARAYER
DLGVVLIIGAMNPLQDCLAPLIGLAGNRYVAVKSPAPATISALIASIDGRVDP
QAVRIYEGSGASARLLECPDHFIFPYSRRRGAQVQSAAMHLPTVLEGGKSPVI
VLEKADRLRAARRIAMKPLINAGQTCVADYILVQEGYVEPLQIGMKRALIYQSDP
```

Query Match 8.4%; Score 63.8; DB 1; Length 11357;

Best Local Similarity 44.0%; Pred. No. 8.6e-08; Mismatches 318; Conservative 0; Mismatches 402; Indels 3; Gaps 1;

```
QY 15 ATTGAAGAGATAMACCTCGTCTTTCGGAGAGAGCTGCAAGAGCATAGCCACATAGG 74
DB 9846 ATCGAAGACACCGCGCTGCGCTTGGCGGCGCGCGCTGCTTGGCGGCGCACATCGG 9787
QY 75 TGTTTGAAGACTATAAGAGCTCGGTATTAAGGTGAGGGCTTTAAGCGGGGTAGCGC 134
DB 9786 CGTGCTCAAGCGAGTAGAGAGACCGGCTTCAGGGCAGAGTGCCTGCGGCGCACAGCAT 9727
QY 135 CGGGGCAATCGTTGGCTTTTATAGCTCAGGCTACTCCCTGAAAGGATGTTGAGCCT 194
DB 9726 CGGTTCTTCATCGCGCGCATGATGCTTTCGCAAGAGCTGCGGAGATCAAGACAGT 9667
QY 195 TCTGAAGAGGTAACTGCTGAGCTGTTTAAGTTCAAGCCACTGTAAGGAGATTGAT 254
DB 9666 CGCGCTCGAGCTGAGCTGCTGATCTGTGCGGGCCCTACCCCTCTCGGCTAGGACTGCT 9607
QY 255 AAGGTGGAGAAAGGCTATAAGATTCTTGAGAAAGTTCTCCTTAAGAGAAATGAAAA 314
DB 9606 CTGCATCCCGCAAGTTGGCAAGATCTCCGCCCGACAGCTCGGTTCCCGCGCATGAGGA 9547
QY 315 ACTTGAGATACGAGGTATATATAGCGGAGAGATTTATCTGCGAAGGCTCTATACCT 374
DB 9546 TGGCGCGCTCTCTCTCGCATGCTGCGCACCGACATCTGCACCGAAGAGAGTGTGCT 9487
QY 375 CTCGAGAGGAGATTAAATCCCGCACTTCTCGGAGCTGAGCAATTCGCGCATATTGA 434
DB 9486 GGGCGAGGGCGACGTGGCCACCGCGGTATGCGCAGCTGCTGATTCGCGGATTTTCAA 9427
QY 435 ACCCGTTGATTAAGATTACTGCTGTTGACGAGGTATATGTTAAACATTTCCGT 494
DB 9426 GCGGCTGAGAGAGGAGATGCTGCTGCTGATGCGGTGATGACCGAAGATGTCGGT 9367
QY 495 TGAAGCCCTTCAGGAAA--GGGGATATCCACCGTTTGGCTGATGCTTCCCATAGA 551
DB 9366 TTCTCGCTCAAGAGATGGGCGGCGGATCGGTGATCGACCTTTTGGCGCGCA 9307
QY 552 GCCGGAAGAGATATTAAGACATTTCTTCAATCTTTTGAAGAGCTTTCTTGCGGT 611
DB 9306 CTCTCTCCCGCGCGGAGAGATCTGTCCGATTTGCTTAAAGCTTTTAAAGCGCAT 9247
QY 612 CGGCTCAAACTCGAAAGAGAGAGTTTGTGACCTGTTATAGTTCTGAGCTTGA 671
DB 9246 GCGGCGCATTCGAGATTCAGATCAGCAAGCGCATGTGTCATCGCTCCCGATCTCTC 9187
QY 672 GGAAGTTCAACCCCTGATGTTAGAAAGAGAGCAAAATATAGAGAGGATACATAA 731
DB 9186 CGGCTCAAGCTGTGATATGTCAGCGCTGCGGAGATTTCGACAGCGGATACAGGA 9127
QY 732 GGC 734
DB 9126 GGC 9124
```

RESULT 3
AE004756 11039 bp DNA linear BCT 30-AUG-2000
LOCUS AE004756 Pseudomonas aeruginosa PA01, section 317 of 529 of the complete
DEFINITION Genome.
ACCESSION AE004756 AE004091

VERSION AE004756.1 GI:9949466
KEYWORDS
SOURCE
ORGANISM Pseudomonas aeruginosa.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.

REFERENCE
1 (bases 1 to 11039)
Stover, C.K., Pham, X.-Q., Erwin, A.L., Mizoguchi, S.D., Warrener, P.,
Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M.,
Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y.,
Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.,
Smith, K., Spencer, D., Wong, G.K., Wu, Z. and Paulsen, I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
Nature 406 (6799), 959-964 (2000)

TITLE

JOURNAL
MEDLINE
PUBMED
10984043
2 (bases 1 to 11039)
Stover, C.K., Pham, X.-Q., Erwin, A.L., Mizoguchi, S.D., Warrener, P.,
Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M.,
Garber, R.L., Goltry, L., Tolentino, E.,
Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N.,
Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,
Wong, G.K.-S., Wu, Z., Paulsen, I.T., Ketzler, J., Sater, M.H.,
Hancock, R.E.W., Lory, S. and Olson, M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA

FEATURES
source
1. 11039
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
90.842
/gene="PA3335"
90.842
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG06723.1"
/db_xref="GI:9949467"
/translation="MSQNDLRTIGTRESAFLERFGRPTITLANPDDVPLLOME
SLMTPEAMVLRCDALVELGVCYGRALLETALINAVLYADLDORALITLRTYIEBG
MSRDPATVDDILNHTRRGASVGRALYLPULGNFRERKLLDSIERSYAAVVS
VPEDSABATRVROSYTRCGVGLBHTRDGIVFTGSDGFSRSYRACLTALALABC
GLTVRSASNLPHACVTLPEGADQGFSSAA"
934.2100
/gene="PA3336"
934.2100
/gene="PA3336"
/codon_start=1
/transl_table=11
/product="probable MFS transporter"
/protein_id="AAG06724.1"
/db_xref="GI:9949468"

gene

CDS

gene

CDS

gene

CDS

complement (2111..3109)
/note="PA3337"
/gene="rfad"
/codon_start=1
/transl_table=11
/product="ADP-L-glycero-D-mannohexose 6-epimerase"
/protein_id="AAG06725.1"

TITLE A Complete Sequence of the *T. tengcongensis* Genome
JOURNAL Genome Res. 12 (5), 689-700 (2002)
MEDLINE 21992816
PUBMED 11997336
REFERENCE 2 (bases 1 to 11618)
AUTHORS Bao, Q., Xu, Z., Hu, S., Dong, W., Chen, Y., Wang, J., Yu, J. and Yang, H.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2001) Beijing Genomics Institute/Genomics and Bioinformatics Center, Institute of Genetics and Development, Chinese Academy of Sciences, Beijing Airport Industrial Zone B6, Beijing 101300, China
REFERENCE 3 (bases 1 to 11618)
AUTHORS Li, W., Xuan, Z., Yang, J., Ling, L. and Chen, R.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2001) Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
REFERENCE 4 (bases 1 to 11618)
AUTHORS Tian, Y., Xue, Y., Xu, Y., Lai, X., Huang, L., Dong, X., Ma, Y. and Tan, H.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2001) Institute of Microbiology, Chinese Academy of Sciences, Beijing 100080, China
FEATURES
source Location/Qualifiers
1. .11618
/organism="Thermoaerobacter tengcongensis"
/strain="WB47"
/db_xref="taxon:119072"
/complement(94. .277)
/gene="TTE1837"
/complement(94. .264)
/gene="TTE1837"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA025028.1"
/db_xref="GI:20516852"
/translation="MECPNCKSTNKGKIGNNLYFCRDCNCEIKIKKCTAVSMYDAEG
CVRKPKVCYNA"
/complement(273. .277)
/gene="TTE1837"
/complement(622. .1485)
/gene="RsaA"
/note="TTE1838"
/complement(622. .1485)
/gene="RsaA"
/note="Best Blastp hit = gi|1664803|sp|O34731|YLBK_BACSU
HYPOTHEICAL.28.3_KDA PROTEIN IN CTAG-RPMF INTERGENIC
REGION gi|7474645|pir||H69874 conserved hypothetical
protein ylbk - Bacillus subtilis
gi|2340007|emb|CAB1357.1| (298682) ylbk protein [Bacillus
subtilis] gi|2633875|emb|CAB1337.1| (299111) similar to
hypothetical proteins [Bacillus subtilis], score 155,
E-value 6.00E-37"
/codon_start=1
/transl_table=11
/product="predicted esterase of the alpha-beta hydrolase
superfamily"
/protein_id="AA025029.1"
/db_xref="GI:20516853"
/translation="MRPRVGLIGGAGARGVHLGLKPEEENIDIPDIVIGSMGAI
IGAIYAGHNVDKLISDAKKINMLKPSLIDFRASRTGLVKGKIKRYLGVYKESPE
ELNPIYIVADIDOTGKGVKEDGLIKAIRASISIPAFEPVINGTLCVDSIYDS
EATIELASLIGADIIINCIVSSIDMGFFERFPLANSEKVLPIKKYFNIRPLPITI
SITTTAKILKDKNSKLEKMEGNKRYITTKPVNDINRWRFQAKCINMGEEAADS
IVARIKRELEK"
/complement(637. .1485)
misc_feature
/gene="RsaA"
/note="Pfam match to entry UPF0028, Uncharacterized
protein family UPF0028, score 39.6, E-value 2.80E-12"
complement(1501. .2084)
/gene="AcrR3"
/note="TTE1839"
complement(1501. .2070)
/gene="AcrR3"

misc_feature
/note="Best Blastp hit = gi|481591|pir||S38906
hypothetical protein 4 - Clostridium pasteurianum
gi|431950|emb|CAA8221.1| (228353) similar to a B. subtilis
gene (GB: BACHEMEHY.5) [Clostridium pasteurianum], score
64.3, E-value 1.00E-09"
/codon_start=1
/transl_table=11
/product="Transcriptional regulator"
/protein_id="AA025030.1"
/db_xref="GI:20516854"
/translation="WESKQSORLAAAYCISTKGYANVSLDIDIEAAYVLSQHY
YPSKKEGLFTREVIKMINIKYLKEINELSGETPAKQKSLVNFPPDLKNNGLPEL
LYDFTGAWMSSESNLSLKDLPNDLSMIEEKLNSALGENDRNYSPPAVAMITGA
MFGQIVILDSNENEILDMLNAQIIFE"
complement(1903. .2043)
/gene="AcrR3"
/note="Pfam match to entry tetr, Bacterial regulatory
proteins, tetr family, score 67.4, E-value 3.00E-16"
complement(2080. .2084)
/gene="AcrR3"
complement(2100. .3788)
/gene="AcrP2"
/note="TTE1840"
complement(2100. .3776)
/gene="AcrP2"
/note="Best Blastp hit = gi|10175329|dbj|BA06427.1|
(AP001516) ABC transporter [Bacillus halodurans], score
373, E-value 1.00E-102"
/codon_start=1
/transl_table=11
/product="predicted unusual protein kinase"
/protein_id="AA025031.1"
/db_xref="GI:20516855"
/translation="WQHLRRYRRLIPVIRYGRGAIIDNIGLKHIVREKIKQND
ENLAKSGERLRLALELPTFRKQIISTRSDDIIPKQITKELEKQDKAPASGD
EVKSVIONEFSESLAEAYAEFPPLAAASIAQVHKLMSGKTVVAVKVRPGIKTI
AQMRLIIEIDIAKPDVNDHTKYGKINFTKRWVEDFKRLSEEDFRIEENAKPEKQNL
KDKVKYPSIITHTTRRLVMEYIGGIPINDFYAIDEAGDRCALRNALKASVLANOI
LRDGFPHGDPHNGINVLBDGTLAFDRGWGSLSPRKQFSGMLGIVKNSRMII
ESIIDNAAVTNANMKLKDINDLRRYVEIPEKXKGVSLNGITDIAVPSYVIVP
NEFMALAKSLITLGGIVKEDPKISVEAFAPIKQILPKFSTQHKKESTINATMYD
SRLIKELPSFLNPLRTEENVAIEKIRPLENLEKRVKVFNRSTSLILLASIV
IAGILIGSGMSANAGAWYKNGIILIKGIAIARVIVGLAISIFRSGRLL"
complement(3784. .3788)
/gene="AcrP2"
complement(3811. .4150)
/gene="TTE1841"
complement(3811. .4137)
/gene="TTE1841"
/note="Best Blastp hit = gi|10175330|dbj|BA06428.1|
(AP001516) BH2709
unknown [Bacillus halodurans], score 70.5, E-value
5.00E-12"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA025032.1"
/db_xref="GI:20516856"
/translation="WTSILERSINFGGLFPLSEKTEKTYVEILVDRCGEVARBDAQK
VMDLVKKEBEKEERLKMIKDAVETGLVNNVAKKEDIVTREIKTSIVREVRVLE
KQNKLEK"
complement(4146. .4150)
/gene="TTE1841"
complement(4233. .4994)
/gene="PncA"
/note="TTE1842"
complement(4233. .4979)
/gene="PncA"
/note="Best Blastp hit = gi|7517111|pir||C72463
hypothetical protein AP02350 - Aeropyrum pernix (strain
KI) gi|5106052|dbj|BA081363.1| (AP000064) 231aa long
hypothetical protein [Aeropyrum pernix], score 60.5,
E-value 2.00E-08"

```
/codon_start=1
/transl_table=11
/product="Amidases related to nicotinamide"
/protein_id="AA025033.1"
/db_xref="GI:20516857"
/translation="MNFDSFLYNTIRPLNYLFDYFSLNLEKSLSSVYNAGBQKVS
IVDMLAGPCKSLGSLAPRAGIIEPIKLLKACYRMGIKNVFLNDAPSDAVFGE
FPPCHVGTSESEIVDELKEIEGEPVIVKSNLNPFGSELEGNEPLKVVEMIKE
KSFIVVGDCTDLVCYQTAMSIKPIANANLKNVIVPENCVETDTSVKTAQSLKI
NPHDNLHITMFLYHMKLNGIEVVKLEEE"
RBS
gene
  complement (5034..5472)
  /gene="PncA"
  complement (5034..5472)
  /gene="Paa12"
  /note="TTTEL843"
  complement (5034..5459)
  /gene="Paa12"
  /note="Best Blastp hit = gi|11499845|ref|NP_071089.1|
  conserved hypothetical protein [Archaeoglobus fulgidus]
  gi|3334444|sp|O28020|YM64_ARCFU_HYPOTHETICAL_PROTEIN
  AF2264 gi|7430238|pir|[H69532 conserved hypothetical
  protein Ar2264 - Archaeoglobus fulgidus
  gi|2648253|gb|AB88986.1| (AE000948) conserved
  hypothetical protein [Archaeoglobus fulgidus], score 87.4,
  E-value 7.00E-17"
  /codon_start=1
  /transl_table=11
  /product="uncharacterized protein, possibly involved in
  aromatic compounds catabolism"
  /protein_id="AA025034.1"
  /db_xref="GI:20516858"
  /translation="MKAINSGIDEKLEPEDLLKVNRTNPHOLLGVHVLGGYATVE
  IEIEKHLNPLNTHAGVLSVMDITMGMAARTVGVQVITINMNYLSPVRVGEVK
  AKGKIHVAGSKTIVVACEYAEADGRLLAVARETFNVLND"
  complement (5043..5405)
  /gene="Paa12"
  /note="Pfam match to entry DUF157, Uncharacterized protein
  Faa1, COG2050, score 103.2, E-value 5.10E-27"
  complement (5468..5472)
  /gene="Paa12"
  misc_feature
  7.3%; Score 55.4; DB 1; Length 11618;
  Best Local Similarity 46.7%; Pred. No. 2.9e-05;
  Matches 216; Conservative 0; Mismatches 241; Indels 6; Gaps 1;
  Qy 26 TAAACCTGTTCTTTCGGAGGAGCTCAAGGCGATAGCCACATAGGTGTTTCAAG 85
  Db 1472 TAGGACTTATCTGGCGGTGGAGCAGAGGATATGCGCACTTAGGTATATAA 1413
  Qy 86 CTATAACGAGCTCGGTATAGGTGAGGCTTTAAGCGGGTGAGCGCGGCAATCG 145
  Db 1412 GATTGAGGAGAGAAATATCTTATGATTTATAGTATAGCATAGTATGGGAGCGAT 1353
  Qy 146 TTTCCGTCCTTTATGCTCAGGCTACTCCCTCAAGGATGTTTCAGCCTTCTCAAGG 205
  Db 1352 TAGGAGCAATCTATGCTTCAGGACATATAGTACAACTTATAGTGATGCAAAA 1293
  Qy 206 TAACTGGCTGAAGCTGTTAAGTTCAAGCCACTCTGAAGGATGATAGGTGGGAGA 265
  Db 1292 TTAATATGTTAAATTTATAAGTTTATGATTTTAAAGCCTCAGCAACCGCTTAG 1233
  Qy 266 AGGCTATAAGATTCCTTGAGGAAGTCTCCCTACAGG-----AGATAGAAACTG 319
  Db 1232 AGGCTGAAATAATAGAGAAATATTTGCGGGTTATGTGAAGAAAGTTTGGAGAACT 1173
  Qy 320 AGATACCGACGTATATGCGCGAGGATTTATCTCGGAGGGCTCTATACCTCTCG 379
  Db 1172 ATATTCCTCTATATTTGTGCTACTGATATCAACAGCAAGGAGTAGTATTTAAG 1113
  Qy 380 AAGGAGTTTAAATCCCGCACTCTCGGAGCTGTGCAATCCCGCATATTTGAACCG 439
  Db 1112 AGGAGACTTGATATAAGCTATAAGGCGAAGCAATTTCTATCCCTGCTTTTGAACCTG 1053
```

Qy 440 TTGAGTATAAGATTACTTCTCTGCTGACGAGGTATAGTTAA 482
Db 1052 TGGAGTACATGAACCAAGTTAGTTGATGATCAATTGTTGA 1010

RESULT 5
ECJ224437
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

ECJ224437.1 GI:3319926
Erwinia carotovora subsp. carotovora expL, expM, and galu genes.
AJ224437
expL gene; expM gene; galu gene; response regulator.
Pectobacterium carotovorum subsp. carotovorum.
Pectobacterium carotovorum subsp. carotovorum
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Pectobacterium.

REFERENCE
1 (bases 1 to 2797)
Andersson, R.A., Palva, E.T. and Pirhonen, M.
The response regulator expM is essential for the virulence of
Erwinia carotovora subsp. carotovora and acts negatively on the
sigma factor RpoS (sigma s)
Mol. Plant Microbe Interact. 12 (7), 575-584 (1999)
99407916
PUBMED
10478478

REFERENCE
2 (bases 1 to 2797)
Andersson, R.A.
Direct Submission
Submitted (13-MAR-1998) Andersson R.A., Plant Biology, Swedish
University of Agricultural Sciences, Box 7080, 750 07 Uppsala,
SWEDEN

FEATURES
source
1..2797
/organism="Pectobacterium carotovorum subsp. carotovorum"
/strain="GCC3193"
/db_xref="taxon:555"
344..1264
/gene="expL"
344..347
/gene="expL"
353..1264
/gene="expL"
/note="rseA homologue"
/codon_start=1
/transl_table=11
/product="ExpL protein"
/protein_id="CAA11941.1"
/db_xref="GI:3319927"
/db_xref="SPTREMBL:O86196"
/translation="MQRKKIGIALGSGAAGKWAHIGVFNALAEIGDIDVDVAGCSIGA
LVGAAYATNNLSSMEQWVRGFGYWDVIRLMDLSWQSGSLRGDRFVNSVHLLHTTQI
EDCAIKYGVVTTNLSTGRELWLTGDLHQAMRASCMPGLLSPYRFNDYVLVDGAVVN
FVPSVLARMAGADIVIAVDLQHDASLNHQDILLSIKPTASEVDIDMEHVSQDWRRIE
RLLRGRQSAESSPTAMEIMNSTSIQILENRLKMTRMAGDDPDVLLQPCQIATLDFH
RAQRAIGAGYKAVAKMRDELLPLATEL"
1370..2397
/gene="expM"
1370..1373
/gene="expM"
1381..2397
/gene="expM"
/function="regulation of virulence factors"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="ExpM protein"
/protein_id="CAA11942.1"
/db_xref="GI:3319928"
/db_xref="SPTREMBL:O86197"
/translation="MEQPLAGKHILVIDEAVFRSVLAGYLTSLGASVREAINGLDAL
SILSHYFEDLMICDLKMPMMGGIEFLERLKSDPTILVISATSMADIAKVLRLGV
QDVLKPIRTLRDVAVMSCLYPMFTSQLNDELQDMQDSLNGSPKATIKLLAQL
QPVQOQTARCRVNYRQLTABQPLVLDIAALSETELAFVCLIDVTGVNNGTAAAL
LLRTLNLGLQELHVDQDQRLPYLPTLLKQVNLRLQASLDGRFLLVGYHYRFLKQL

IIISAGIANTATVNEQVIVLNSGVPLGTILEGATVLAQVLYOCETWOCQIIGGGGRLLRLM
LTTTB"
gene 2593 .2797
/gene="galu"
RBS 2593 .2597
/gene="galu"
CDS 2604 .>2797
/gene="galu"
/codon_start=1
/transl_table=11
/product="GalU protein"
/protein_id="CAA11943.1"
/db_xref="GI:3319929"
/translation="MSIVKVKVKAIVIPVAGIGTSMPLPATKAIPKEMLPLVDEKPLIYQ
VNECIAAGINIIIVTHSS"
BASE COUNT 703 a 562 c 756 g 776 t
ORIGIN
Query Match 7.3%; Score 55.2; DB 1; Length 2797;
Best Local Similarity 54.4%; Pred. No. 3.2e-05;
Matches 111; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
QY 341 GCACGATTATATACCGGAGAGGCTCTATACCTCTCCGAAAGGAGTTATCCCGCAC 400
DB 690 GCACCAATCTGACGACGGGGCGTGAACCTGCTGCTACAGAAAGCGATTACACAGGCA 749
QY 401 TTCTCGGAGCTGTGCATTTCCCGGCAATTTGAACCCGTTGAGATATAGATTAATCTGC 460
DB 750 TCGGGGCGCTTTCGAGTATGCTGCTGCTTTATCCCGCTTGATTAACATTAATCTGC 809
QY 461 TCGTTGACGAGAGTATGATTAAACCTTCCGTTAGACCTTTGAGAAAGCGTATTC 520
DB 810 TGGTGGACGGCGCGGGGTGTTATCTGTTCCGTTTCTGCGCGGCAATGGGGCTG 869
QY 521 CCACGGTTGGCTGATGTCCTTC 544
DB 870 ATATGCTATGTCGTATATCTCC 893
RESULT 6
SME591787/c 329100 bp DNA linear BCT 05-JUN-2002
LOCUS SME591787 329100 bp DNA linear BCT 05-JUN-2002
DEFINITION Sinorhizobium meliloti 1021 complete chromosome; segment 6/12.
ACCESSION AL591787 AL591688
VERSION AL591787.1 GI:15074266
KEYWORDS
SOURCE
ORGANISM Sinorhizobium meliloti.
Sinorhizobium meliloti
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Sinorhizobium.
REFERENCE
AUTHORS 1 (bases 1 to 329100)
Capela, D., Barloy-Hubler, F., Gouzy, J., Bothe, G., Ampe, F., Batut, J.,
Boisnard, P., Becker, A., Boutry, M., Cadieu, E., Dreano, S., Gloux, S.,
Godrie, T., Goffeau, A., Kahn, D., Kiss, E., Lelaure, V., Maury, D.,
Pohl, T., Portetelle, D., Puehler, A., Purnelle, B., Rameberger, U.,
Renard, C., Thebaud, P., Vandenbol, M., Weidner, S. and Galibert, F.
Analysis of the chromosome sequence of the legume symbiont
Sinorhizobium meliloti strain 1021
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)
JOURNAL MEDLINE 2136507
MEDLINE 11481430
PUBMED 2 (bases 1 to 329100)
REFERENCE Gouzy, J.
AUTHORS Direct Submission
JOURNAL Submitted (26-JUN-2001) Gouzy J., Submitted on behalf of the MELILO
COMMENT
MELILO EU Consortium:
Laboratoire de Biologie Molculaire des Relations
Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
France, Laboratoire de Genetique et Developpement UMR6061-CNRS,
Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,
France, GARC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,

Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,
D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,
Universite Catholique de Louvain, Place Croix du Sud 2, Bre 20,
B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte
des Sciences Agronomiques de Gembloux, Avenue Maredchal Juin 6,
B-5030 Gembloux, Belgium. E-mail:jerome.gouzy@oulouise.inra.fr
http://sequence.coulouise.inra.fr/meliloti.html.
location/Qualifiers
1. 329100
/organism="Sinorhizobium meliloti"
/strain="1021"
/DB_xref="taxon:382"
152. .227
/gene="trna-thr_tgt"
152. .227
/gene="trna-thr_tgt"
/product="trna-thr_tgt"
/note="codon recognized: ACA; predicted by tRNAscan-SE"
/evidence="not experimental"
complement (373. .798)
/gene="SMC04435"
complement (373. .798)
/gene="SMC04435"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence: hypothetical"
Gene name confidence: hypothetical
predicted by Homology"
/codon_start=1
/evidence="not experimental"
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAC45913.1"
/db_xref="GI:15074267"
/db_xref="SPTREMBL:Q92019"
/translation="MNALVPEPFAVDWQASKEFCNIIIGFSCDYEPPEEGCYSLD
GALMLIDQIGTGTADGHLPRSPYGLGNTVQIRVASADALVALTDGIALYLPLE
ERWYRRGDEVEGRCFVADPDGYLRFYEPGRPRRTS"
complement (914. .1855)
/gene="SMC01330"
complement (914. .1855)
/gene="SMC01330"
/function="miscellaneous; not classified regulator"
/note="Product confidence: putative"
Gene name confidence: hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence="not experimental"
/transl_table=11
/product="putative TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAC45914.1"
/db_xref="GI:15074268"
/db_xref="SPTREMBL:Q92018"
/translation="MODLNDIALPAVVGNGFSAARDLVNPKSLKSHVARLEBOL
GVRLERSRTRKRMTEVGRIFVEHOGGLDGVAAAEATIAVARPTGVPLACIGF
TMDLADITPAFRRRRVGYVLLTASRRRIDLEBIDVALRKDDLDTSQILVAKFG
EVRQDLIAERLARGITTDNISQMPFLSNBQHPNDVWRLVHATGDALEIARPV
VGSDFLIBERAIKRGMGIALPDHICBAFPTGALVPTLPMTSGNVVHVLVPSRH
GLPATRALIDFLAENTLIKALERCHEVDPRAASFEI"
2063. .2683
/gene="acpD OR SMC01329"
2063. .2683
/gene="acpD OR SMC01329"
/function="small molecule metabolism; fatty acid
biosynthesis"
/note="Product confidence: putative"
Gene name confidence: putative
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence="not experimental"

```

/transl_table=11
/product="PROBABLE ACYL CARRIER PROTEIN PHOSPHOTESTERASE"
/protein_id="CAC45915.1"
/db_xref="GI:15074269"
/db_xref="SPTREMBL:Q92017"
/translation="MKILHDSGILGSHSVSRRLTAIVSOLKADRPDAEITYRDLAS
ERVPHLTAQIMAPADLEGVADLAADVRIGROMLREFLAADVVGAPWYFIPSQ
LKAWIDRLAVAGTKFRVTEAGPEGLAKGKLIIVASTRGHYSVAPASAMDHQETYLRS
VEGPGGTDIEFTIRAGLNLGPOKQPAIAEABKTAEGNVLKAS"
2792..3076
/gene="SMC01328"
CDS
2792..3076
/gene="SMC01328"
/function="miscellaneous; hypothetical/partial homology"
/translation="MAKSEPERKPHVNIIGTIGHVDHGKTSLSAAITKYFGFKNAYDQ
IDAAPERKAGITITSTARVEYETNRHYHVDVCFGHADYVKNMTTGAQMDGALLVVS
AADGPMQTRREHILLARQVPAIVFLNKVDQVDDALLELELLELVRELLSYEPFG
DDIPVKSALAEALSDSKKIGEDAIRLMAADVAYIPTPERPIDQPLMPLEIDVFSI
SGRTVTVTRVERGIKVGGEIEIVGIRPTTKTCTGVEMPRKLLDQAGADNIGALL
RGVRNGVERQIILCKFGSVKPRKFAEAYILTKESGGRTTPTTINRPOFYFRITD
VTGIVTLPEGTEMVMPGDNVTVDVELIVPIAMEEKLRFAIREGRTVAGIVASIV"
6125..6790
/gene="SMC01325"
CDS
6125..6790

Query Match 7.3%; Score 55; DB 1; Length 329100;
Best Local Similarity 46.0%; Pred. No. 4.4e-05;
Matches 227; Conservative 0; Mismatches 260; Indels 6; Gaps 1;

QY 24 GATAAACCTCGTCTTTTCGGGAGGAGCTGCAAGAGGATACCCACATAGGTGTTTGA 83
DB 78470 GATTGCGTGGCGTTCGGCGCGGTGCTGCGCGGGCTGGGCCCATATCGGTGCTGCTG 78411

QY 84 AGCTATAAACAGAGCTCGGTATAGGGGTGAGGCGCTTTAAGCGGGGTGAGCGCGCGGCAAT 143
DB 78410 AGCCCTCGACGAGGAGGATATCGAAGTCGGAATGATTGCGGGCACATCGATCGGTGCGCT 78351

QY 144 CGTTTCGTCTTTTATGCTCAGGCTACTCCCTTGAAGGGAGTGT-----CAGCCTTCT 197
DB 78350 GGTGGCGGCTGTTTACCTCGCGGCAAGCTCGATGAATTTGAGGCTTTTCGCCCGCTCGCT 78291

QY 198 GAAGAGGTAACCTGGCTGAAGCTGTTTAAAGTTCACGCCACCTCTGAAGGATGATAGG 257
DB 78290 TACCGTGGCGCGATTTGCGGCTTTCGATTTTCGGATCGCGCGGTGCGCTTCTCGG 78231

QY 258 GTGGAGAGAGCTATAAGATTCTTTGAGGAAGTTCTCCCTTTACAGGAGATAGAAAAACT 317
DB 78230 CGGTCTGGCGCTCACCAGCGGATGCGAGAACATCTGCAGAACCTCAGTATCAGGATCT 78171

QY 318 TGAGATACCGAGCTATATATGCGGACGAGATTATATCTCGGGAAGGGCTCTATACCTCTC 377
DB 78170 CGACCGGCGCTTCTGTCGCGCTGCGCCAGGAGTCTACAGCGGTACAGAGGTATGATCGA 78111

QY 378 GGAAGGAGATTAAATCCCGCAGCTTCTCGGCGAGCTGTCGAATTCGCCGATATTTGAACC 437
DB 78110 GAAAGGATCCCTTATACCGCATTTGCGCATCTTACGCTTGTGCTGCGCATTTTCGAGCC 78051

QY 438 CGTTGAGTATAAGAAATTACTTGTCTGTTGACGGAGGTATAGTTAAACACCTTCCCGTTGA 497
DB 78050 CGTCAACGCCAATGGCGGACTCTCATCGACGCGCGCTGTGTCAACCCCGTTCGCGTTTC 77991

QY 498 GCCCTTTCAGGA 510
DB 77990 GGTCTGCGCGCA 77978

RESULT 7
A96528
LOCUS A96528 903 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 561 from Patent WO924578.
ACCESSION A96528
VERSION A96528.1 GI:6780156
KEYWORDS
SOURCE unidentified.

```

```

/transl_table=11
/product="PROBABLE ACYL CARRIER PROTEIN PHOSPHOTESTERASE"
/protein_id="CAC45915.1"
/db_xref="GI:15074269"
/db_xref="SPTREMBL:Q92017"
/translation="MKILHDSGILGSHSVSRRLTAIVSOLKADRPDAEITYRDLAS
ERVPHLTAQIMAPADLEGVADLAADVRIGROMLREFLAADVVGAPWYFIPSQ
LKAWIDRLAVAGTKFRVTEAGPEGLAKGKLIIVASTRGHYSVAPASAMDHQETYLRS
VEGPGGTDIEFTIRAGLNLGPOKQPAIAEABKTAEGNVLKAS"
2792..3076
/gene="SMC01328"
CDS
2792..3076
/gene="SMC01328"
/function="miscellaneous; hypothetical/partial homology"
/translation="MAKSEPERKPHVNIIGTIGHVDHGKTSLSAAITKYFGFKNAYDQ
IDAAPERKAGITITSTARVEYETNRHYHVDVCFGHADYVKNMTTGAQMDGALLVVS
AADGPMQTRREHILLARQVPAIVFLNKVDQVDDALLELELLELVRELLSYEPFG
DDIPVKSALAEALSDSKKIGEDAIRLMAADVAYIPTPERPIDQPLMPLEIDVFSI
SGRTVTVTRVERGIKVGGEIEIVGIRPTTKTCTGVEMPRKLLDQAGADNIGALL
RGVRNGVERQIILCKFGSVKPRKFAEAYILTKESGGRTTPTTINRPOFYFRITD
VTGIVTLPEGTEMVMPGDNVTVDVELIVPIAMEEKLRFAIREGRTVAGIVASIV"
6125..6790
/gene="SMC01325"
CDS
6125..6790

Query Match 7.3%; Score 55; DB 1; Length 329100;
Best Local Similarity 46.0%; Pred. No. 4.4e-05;
Matches 227; Conservative 0; Mismatches 260; Indels 6; Gaps 1;

QY 24 GATAAACCTCGTCTTTTCGGGAGGAGCTGCAAGAGGATACCCACATAGGTGTTTGA 83
DB 78470 GATTGCGTGGCGTTCGGCGCGGTGCTGCGCGGGCTGGGCCCATATCGGTGCTGCTG 78411

QY 84 AGCTATAAACAGAGCTCGGTATAGGGGTGAGGCGCTTTAAGCGGGGTGAGCGCGCGGCAAT 143
DB 78410 AGCCCTCGACGAGGAGGATATCGAAGTCGGAATGATTGCGGGCACATCGATCGGTGCGCT 78351

QY 144 CGTTTCGTCTTTTATGCTCAGGCTACTCCCTTGAAGGGAGTGT-----CAGCCTTCT 197
DB 78350 GGTGGCGGCTGTTTACCTCGCGGCAAGCTCGATGAATTTGAGGCTTTTCGCCCGCTCGCT 78291

QY 198 GAAGAGGTAACCTGGCTGAAGCTGTTTAAAGTTCACGCCACCTCTGAAGGATGATAGG 257
DB 78290 TACCGTGGCGCGATTTGCGGCTTTCGATTTTCGGATCGCGCGGTGCGCTTCTCGG 78231

QY 258 GTGGAGAGAGCTATAAGATTCTTTGAGGAAGTTCTCCCTTTACAGGAGATAGAAAAACT 317
DB 78230 CGGTCTGGCGCTCACCAGCGGATGCGAGAACATCTGCAGAACCTCAGTATCAGGATCT 78171

QY 318 TGAGATACCGAGCTATATATGCGGACGAGATTATATCTCGGGAAGGGCTCTATACCTCTC 377
DB 78170 CGACCGGCGCTTCTGTCGCGCTGCGCCAGGAGTCTACAGCGGTACAGAGGTATGATCGA 78111

QY 378 GGAAGGAGATTAAATCCCGCAGCTTCTCGGCGAGCTGTCGAATTCGCCGATATTTGAACC 437
DB 78110 GAAAGGATCCCTTATACCGCATTTGCGCATCTTACGCTTGTGCTGCGCATTTTCGAGCC 78051

QY 438 CGTTGAGTATAAGAAATTACTTGTCTGTTGACGGAGGTATAGTTAAACACCTTCCCGTTGA 497
DB 78050 CGTCAACGCCAATGGCGGACTCTCATCGACGCGCGCTGTGTCAACCCCGTTCGCGTTTC 77991

QY 498 GCCCTTTCAGGA 510
DB 77990 GGTCTGCGCGCA 77978

RESULT 7
A96528
LOCUS A96528 903 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 561 from Patent WO924578.
ACCESSION A96528
VERSION A96528.1 GI:6780156
KEYWORDS
SOURCE unidentified.

```


RESULT 9
 A96526 903 bp DNA linear PAT 07-SBP-2000
 LOCUS Sequence 559 from Patent WO9924578.
 DEFINITION A96526
 ACCESSION A96526
 VERSION A96526.1 GI:6780155
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE
 1 (bases 1 to 903)
 AUTHORS Pizze, M., Scariato, V., Rappuoli, R., Grandi, G. and Masignani, V.
 TITLES Neisserial antigens
 JOURNAL Patent: WO 9924578-A 559 20-MAY-1999;
 PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
 CHIRON SPA (IT); GRANDI GUIDO (IT); MASIGNANI VEGA (IT)
 FEATURES
 source location/Qualifiers
 1. 903
 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 204 a 230 c 255 g 214 t
 ORIGIN
 Query Match 6.9%; Score 52; DB 6; Length 903;
 Best Local Similarity 44.3%; Pired. No. 0.00028;
 Matches 211; Conservative 0; Mismatches 265; Indels 0; Gaps 0;
 Oy 19 GAGAGATTAACCTCGTTCTTTGCGGAGAGCTGCGAAGGCGATGACCATAGATGTT 78
 Db 130 GCGGTGTCGTTTGGCACTCGGTGGCGCGCATTAAGATTTGCCATGATATT 189
 Oy 79 TTGAAGCTATAACAGAGCTCGGTATAGGGGTAGGGCTTTAAGCGGGGTAGCGCGG 138
 Db 190 ATTACGTTTGAAGAAAGGATATTCCTGTAAGGTGTTACCGGCACATCGCAGGT 249
 Oy 139 GCAATGTTTGGTCTTTTATGCTTCAAGCTACTCCCTGAAAGGATGTTCAAGCTTCG 198
 Db 250 TCGATGTGCGGAGCGCTTTTGGATCGGGATGTCGCCGACCGCTCGAATTTGGAAGC 309
 Oy 199 AAGAGGTAACTGCTGGAAGCTGTTAAGTTCAAGCCACTGTGAAGGATTTGATGG 258
 Db 310 GAAATTTTGGGCAAAACGATTTGTCGATTTAACTTTCACCGAGTGGTTTATCAA 369
 Oy 259 TGGAGAAAGCTTATGATTTCTTGAAGAGTTCTCCCTTAACGAGATTAAGAAACTT 318
 Db 370 GCGGAAAGCTGCAAAATTCATCAACGAAAGTGGCGGCGAGATTCAGCATTT 429
 Oy 319 GAGATACCGAGATATATGCGGAGCGATTTACTCGGGAAGGCTCTATACCTCTCG 378
 Db 430 CCCATCAAAATTTCCGCGCGTGTCTACGATTTTGAACCGGGAAGCGCTCGCTTCAT 489
 Oy 379 GAAGGAGTTTATCCCGGCACTTCTCGGAGCTGTGCAATCCCGGCAATTTGAACC 438
 Db 490 CAGGGGAATGCGCGGAGGCTGTGCGCGCTTCCCGCCCATTTCCCAATGTGTTCAACC 549
 Oy 439 GTTGAATTAAGATTAATCTGCTGTTGACGAGGATTAATTAACAACCTCCCGT 494
 Db 550 GTTATCATCGGCGAGGATCATATGTTGACGCGCGGTCTGCGACGCCCGTCCCGT 605

RESULT 10
 AE002562/ 9831 bp DNA linear BCT 25-MAY-2000
 LOCUS Neisseria meningitidis serogroup B strain MC58 section 204 of 206
 DEFINITION of the complete genome.
 ACCESSION AE002562 AE002098
 VERSION AE002562.1 GI:7227392
 KEYWORDS
 SOURCE Neisseria meningitidis MC58.
 ORGANISM Neisseria meningitidis MC58.
 Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
 Neisseria.
 REFERENCE
 1 (bases 1 to 9831)

AUTHORS
 Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C.,
 Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F.,
 Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D.,
 Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D.,
 Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E.,
 Clifton, H., Clark, E.B., Cotton, M.D., Uterback, T.R., Khouri, H.,
 Qin, H., Vamathevan, J., Gill, J., Scariato, V., Masignani, V.,
 Pizze, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,
 Rappuoli, R. and Venter, J.C.
 TITLE
 Complete genome sequence of Neisseria meningitidis serogroup B
 strain MC58
 JOURNAL
 Science 287 (5459), 1809-1815 (2000)
 MEDLINE
 20175755
 PUBMED
 10710307
 REFERENCE
 2 (bases 1 to 9831)
 AUTHORS Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C.,
 Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F.,
 Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D.,
 Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D.,
 Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E.,
 Clifton, H., Clark, E.B., Cotton, M.D., Uterback, T.R., Khouri, H.,
 Qin, H., Vamathevan, J., Gill, J., Scariato, V., Masignani, V.,
 Pizze, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,
 Rappuoli, R. and Venter, J.C.
 TITLE
 Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
 Direct Submission
 Medical Center Dr, Rockville, MD 20850, USA
 FEATURES
 source location/Qualifiers
 1. 9831
 /organism="Neisseria meningitidis MC58"
 /strain="MC58"
 /db_xref="taxon:122586"
 /note="serogroup: B"
 772. 2229
 /gene="NMB2136"
 /note="similar to GB:AL009126 percent identity: 59.41;
 identified by sequence similarity; putative"
 /codon_start=1
 /product="peptide transporter"
 /protein_id="AAF2444.1"
 /db_xref="GI:7227393"
 /translation="MSRHPAPYGEKTFRFPOLSTFHELMERSFYMGQILITY
 LVTADGSGVIDKTLNAGIVAGVSGSVYSTLIGAFVADRWGAERTPLDSIGVNL
 GHVLAAPGIVGELIGLIFIALGSGGVKSTASMGVALYEDKMRPLRPDAGSIVPI
 AINIGFPLGDLVGLQENIGFHYFGAAVAGMAFGIMRSLGRKNLPHPTVPHLSK
 GCGKTAAVGIALIALATAIKTGLVNLDFSGILSTVLAIVAPARLLTPRVSS
 DNKRHTIAYTPIFLTCMPVAWVFOITVATVTFEDVFNVTSSFPYPAVMKMSGL
 WILFSGIIMAWMTKMRKOPKTPKRAVAAPVTGASFLDFVPFISGTPMPVAVPL
 IVLITIGELMISPLAISITKIAPLPFTQWALNPLAFSLGFTLGGLVFERGYONG
 DEIGFYRLFLFYIGATGTFLLLLVPKLNKLMBGTD"
 2752. 2964
 /gene="NMB2137"
 2752. 2964
 /gene="NMB2137"
 /note="hypothetical protein; identified by Glimmer2;
 putative"
 /codon_start=1
 /trnasl_table=11
 /product="hypothetical protein"
 /protein_id="AAF2445.1"
 /db_xref="GI:7227394"
 /translation="MLKQPTRTLGGGRHCVLFGIAPNGVWPAAYCYQMRGAPLPHLFT
 LTCAKAALIGGPAFCSTFRVTAAGR"
 complement (3122. 4425)
 /gene="NMB2138"
 complement (3122. 4425)
 /gene="NMB2138"
 /note="similar to GB:M38590 SB:P28353 PID:154276 percent
 identity: 80.11; identified by sequence similarity;
 putative"

Db 4741 CAGGGGAATCCGGGAGGCTGTGGCGCTTCGCCGCCATTCCTCCCAATGTGTTCCACCC 4682
 Qy 439 GTTGAATATAGAAATTAATCTGCTGTTGACGAGATATAGTTAAACAATTCCTCCGT 494
 4681 GTTATCATCGGAGGAGATACATATATGTGACGCGGTCTGTCCGACCCCTGCCCCGT 4626

RESULT 11
 AX044035 172325 bp DNA linear PAT 24-NOV-2000
 LOCUS AX044035/c
 DEFINITION Sequence 114 from Patent WO0066791.
 ACCESSION AX044035
 VERSION AX044035.1 GI:11342919
 KEYWORDS
 SOURCE
 ORGANISM
 Neisseria meningitidis.
 Neisseria meningitidis
 Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
 Neisseria.
 1 (bases 1 to 172325)
 Pizza, M., Hickey, E., Peterson, J., Tettelin, H., Venter, J. C.,
 Massignani, V., Galeotti, C., Mora, M., Ratti, G., Scarselli, M.,
 Scarlato, V., Rappelli, R., Frazer, C. M. and Grandi, G.
 Neisseria genomic sequences and methods of their use
 Patent: WO 0066791-A 114 03-NOV-2000;
 JOURNAL CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
 TITLE Location/Qualifiers
 FEATURES
 source
 1. 172325
 /organism="Neisseria meningitidis"
 /db_xref="taxon:487"
 /note="sequence too long, cut in 8 pieces.-seq 1: 1 to
 349980 349980 bases-seq 108: 300001 to 649980 349980
 bases-seq 109: 600001 to 949980 349980 bases-seq 110:
 900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980
 349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq
 113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to
 2272325 172325 bases"

BASE COUNT 43072 a 47583 c 41465 g 40205 t
 ORIGIN

Query Match 6.9%; Score 52; DB 6; Length 172325;
 Best Local Similarity 44.3%; Pred. No. 0.00034;
 Matches 211; Conservative 0; Mismatches 265; Indels 0; Gaps 0;

Qy 19 GAAGAGATAAACCCTGTTCTTCCGAGAGAGCTGCAAGGCGATAGCCCATAGGTCT 78
 Db 153800 GGAAGGTCGCTTGGCACTCGGTGGCGCGCATTTAAAGATTGGCCCATGTAGATT 153741

Qy 79 TTGAAGCTATAACGAGCTCGTATTAAGGGTGAGGGCTTTAAGCGGGATAGCCCGG 138
 Db 153740 ATTAAGTATTAAGAAAGAAACGATATTCCTGTGAAGTGTTACCGGCACATCGCAGGT 153681

Qy 139 GCAATCGTTGGCTTTTATAGCCCTCAGGCTACTCCCTGAAGGATGTACGCTTGTG 198
 Db 153680 TCGATTGTGGCAGGCTTTTTCATCGGTATGTGCCGACCGCTGAATTGGAAACC 153621

Qy 199 AAGAGGTAACCTGCTGAAGCTGTTTAAGTTCAAGCACCTGTGAAGGATATGATAGG 258
 Db 153620 GAATTTTAAAGCAAAACGATTTGTGTCAATTAACCTGTCCACAGTGCTTTATCAAA 153561

Qy 259 TGGGAGAGGCTATAAGATTCCTTGAGAGAGTTCTCCCTTACAGAGAAATGAAGAACTT 318
 Db 153560 GCGGAAAGCTGCAAAATTAATCAACGAAAGTGGCGGACAGGACAGATTCAGAGTTT 153501

Qy 319 GAGATACGAGGTATATATNGCGGAGATTTATCTCGGGAAGGGCTCTATACCTCTCG 378
 Db 153500 CCCATCAATTTTCCCGCGGTGTCTACTATATTTTGAACCGCGAAGCGCTTTCAT 153441

Qy 379 GAAGGAGTTTATATCCCGCATCTTCGCGAGCTGTGCAATTCGCGGATATTTGAAGCC 438
 Db 153440 CAGGGGATGCCCGGACAGGCTGTGCGGCTTCGCGGCATTTCCAAAGTGTTCACACC 153381

Qy 439 GTTGAATATAGAAATTAATCTGCTGTTGACGAGATATAGTTAAACAATTCCTCCGT 494

Db 153380 GTTATCATCGGAGGAGATACATATATGTTGACGGGCTGTGCGACGCCGCT 153325

RESULT 12
 AP001516/c 300950 bp DNA linear BCT 10-JAN-2001
 LOCUS AP001516
 DEFINITION Bacillus halodurans genomic DNA, section 10/14.
 ACCESSION AP001516 BA000004
 VERSION AP001516.1 GI:10175192
 KEYWORDS
 SOURCE
 ORGANISM
 Bacillus halodurans DNA.
 Bacillus halodurans
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 1 (sites)
 Takami, H., Nakasone, K., Hirama, C., Takaki, Y., Masui, N., Fujii, F.,
 Nakamura, Y. and Inoue, A.
 An improved physical and genetic map of the genome of alkaliphilic
 Bacillus sp. C-125
 Extremophiles 3 (1), 21-28 (1999)
 99184645
 10086841
 JOURNAL MEDLINE
 PUBMED
 10086841
 REFERENCE
 AUTHORS
 Takami, H. and Horikoshi, K.
 IDENTIFICATION of facultatively alkaliphilic Bacillus sp. C-125
 to Bacillus halodurans
 Biosci. Biotechnol. Biochem. 63, 943-945 (1999)
 3 (sites)
 Takami, H., Nakasone, K., Ogasawara, N., Hirama, C., Nakamura, Y.,
 Masui, N., Fujii, F., Takaki, Y., Inoue, A. and Horikoshi, K.
 Sequencing of three lambda clones from the genome of alkaliphilic
 Bacillus sp. strain C-125
 Extremophiles 3 (1), 29-34 (1999)
 99184645
 10086842
 JOURNAL MEDLINE
 PUBMED
 10086842
 REFERENCE
 AUTHORS
 Takami, H., Takaki, Y., Nakasone, K., Hirama, C., Inoue, A. and
 Horikoshi, K.
 Sequence analysis of a 32-kb region including the major ribosomal
 protein gene clusters from alkaliphilic Bacillus sp. strain C-125
 Biosci. Biotechnol. Biochem. 63 (2), 452-455 (1999)
 99209008
 10192928
 JOURNAL MEDLINE
 PUBMED
 10192928
 REFERENCE
 AUTHORS
 Takami, H., Takaki, Y., Nakasone, K., Sakiyama, T., Maeno, G.,
 Sasaki, R., Hirama, C., Fujii, F. and Masui, N.
 Genetic analysis of the chromosome of alkaliphilic Bacillus
 halodurans C-125
 Extremophiles 3 (3), 227-233 (1999)
 99411980
 10484179
 JOURNAL MEDLINE
 PUBMED
 10484179
 REFERENCE
 AUTHORS
 Takami, H.
 Genome analysis of facultatively alkaliphilic Bacillus halodurans
 C-125
 (in) Extremophiles in deep-sea environments (Ed.);
 HORIKOSHI, K. TSUJII;
 : 249-284; Springer-Verlag (1999)
 7 (sites)
 Takami, H., Masui, N., Nakasone, K. and Horikoshi, K.
 Replication origin region of the chromosome of alkaliphilic
 Bacillus halodurans C-125
 Biosci. Biotechnol. Biochem. 63 (6), 1134-1137 (1999)
 99356711
 10427704
 JOURNAL MEDLINE
 PUBMED
 10427704
 REFERENCE
 AUTHORS
 Takami, H. and Horikoshi, K.
 Analysis of the genome of an alkaliphilic Bacillus strain from an
 industrial point of view
 Extremophiles 4 (2), 99-108 (2000)
 20263314
 10805564
 JOURNAL MEDLINE
 PUBMED
 10805564
 REFERENCE
 9 (sites)

AUTHORS Nakasone, K., Masui, N., Takaki, Y., Sasaki, R., Maeno, G., Sasaki, R., Sakiyama, T., Hirama, C., Fuji, P. and Takami, H.
TITLE Characterization and comparative study of the rrm operons of alkaliphilic *Bacillus halodurans* C-125
JOURNAL Extremophiles 4 (4), 209-214 (2000)
MEDLINE 20426005
PUBMED 10972189
REFERENCE 10 (sites)
AUTHORS Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., Fuji, P., Hirama, C., Nakamura, Y., Ogasawara, N., Kuhara, S. and Horikoshi, K.
TITLE Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*
JOURNAL Nucleic Acids Res. 28 (21), 4317-4331 (2000)
MEDLINE 20512582
PUBMED 11058132
REFERENCE 11 (bases 1 to 300950)
AUTHORS Takami, H. and Takaki, Y.
TITLE Direct Submission
JOURNAL Submitted (22-MAR-2000) Hideto Takami, Japan Marine Science and Technology Center, Deep-sea Microorganisms Research Group; 2-15 Natsushima, Yokosuka, Kanagawa 237-0061, Japan
MEDLINE (E-mail: takami@jamstec.go.jp)
JOURNAL URL: <http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/Fresearch.html>,
 Tel: 81-468-67-3895, Fax: 81-468-66-6364

FEATURES
 source

gene
 CDS
 /organism="Bacillus halodurans"
 /db_xref="taxon:86665"
 /note="alkaliphile"
 /complement(179..2113)
 /gene="spoVD"
 /complement(179..2113)
 /gene="spoVD"
 /note="BH2572"
 /codon_start=1
 /transl_table=11
 /product="stage V sporulation protein (sporulation specific penicillin-binding protein) (spore cortex)"
 /protein_id="BAB06291.1"
 /db_xref="GI:10175193"
 /translation="MRVSNIVTNRRLIFVLLVGLATFLIIMRLGVVOPALGDMLTQ
 AODSWSDVPEAKRGEITDRNDVLATNAPSILVPRVQKDAETAKLANVLQ
 DRKAYELIKRESIVLNPGBKISKDKANVRLKRLPGVIAEDSKHVPKGSLS
 HVLGAGIDQGLTGLYDQQLKEGSHVSPDQKQRLPNLADSTAPINGLNL
 FLITDSRVITIRELDIAETISPDGATAINPNPTGILGMSRPHYNPNFREP
 PEVYNQKPIIMVYEPGTFKIITLAALEENVEDITDRFDPGFIEVAGHLECW
 KGGHSGTFLEWVNSCNPFGVILGERLGRDLYIEAFGQKGTGLDQGEAKGII
 FNRDRIGLEQATFAGQSVSTPIQVAAVSAVNGVLYEPYAKDWDPYTGCVY
 ESSAPKMKRQIISBETSKVRYALESVWARGTQCGAFVGYRVGGKTGTAOKAGRY
 LNNHIVSFIQFADDPQIIVVVAIDNPKDTVQFGVVAAPIVGKIIGDSIQALGIE
 KRNQIEKERKNDPELVEVDLVGRTRKDLHESYYELKIDADGKGVVVAQAPERT
 KVAVGSTIRLYMDDTPTD"
 /complement(2295..4466)
 /gene="pbbp"
 /complement(2295..4466)
 /gene="pbbp"
 /note="BH2573"
 /codon_start=1
 /transl_table=11
 /product="penicillin-binding protein 2B (cell-division septum)"
 /protein_id="BAB06292.1"
 /db_xref="GI:10175194"
 /translation="MDIKRAATNKRAMLVLFIVVFSVLIAFLYLQIEKVKYSL
 QAIABARTSVSEVLHKGRTIRDRGDALAEVTSVVAVLNDSHOPAVEPRTAQ
 LLAHINNSERLLELTDRVQVGLGAGANLTQQRDKIADLDLPGIYFTEERY
 YPKHYASHLIGITDRDMGSRGLESDEYLRGDSRIRFKQCGIIPRPSERI
 EPKATGNDVYITLDSNIQTALDQVMTQVEEYKPERIIAIVADPKTQGLIAMSPTF
 NPNEYQITNTMYAVSDRPGSTMKVFTLAAALEEVLNVEQYQSGTYAIRNDTN
 PPIRDHQCRGWGITISYLEAMQSSNVGFSIALEKLGPEKLYLIDFEGFGEITD
 LPNEAASVYKALRDAASTAFQGTAVTPIQIQIQAATAIANDGKMKPVVVDRTVDS
 ETGEVIEKEPEVVGEPISEKETAKEVRDILETVTVTSSSGTGRPFYLEGPDVAGTGA

gene
 CDS
 /organism="Bacillus halodurans"
 /db_xref="taxon:86665"
 /note="alkaliphile"
 /complement(179..2113)
 /gene="spoVD"
 /complement(179..2113)
 /gene="spoVD"
 /note="BH2572"
 /codon_start=1
 /transl_table=11
 /product="stage V sporulation protein (sporulation specific penicillin-binding protein) (spore cortex)"
 /protein_id="BAB06291.1"
 /db_xref="GI:10175193"
 /translation="MRVSNIVTNRRLIFVLLVGLATFLIIMRLGVVOPALGDMLTQ
 AODSWSDVPEAKRGEITDRNDVLATNAPSILVPRVQKDAETAKLANVLQ
 DRKAYELIKRESIVLNPGBKISKDKANVRLKRLPGVIAEDSKHVPKGSLS
 HVLGAGIDQGLTGLYDQQLKEGSHVSPDQKQRLPNLADSTAPINGLNL
 FLITDSRVITIRELDIAETISPDGATAINPNPTGILGMSRPHYNPNFREP
 PEVYNQKPIIMVYEPGTFKIITLAALEENVEDITDRFDPGFIEVAGHLECW
 KGGHSGTFLEWVNSCNPFGVILGERLGRDLYIEAFGQKGTGLDQGEAKGII
 FNRDRIGLEQATFAGQSVSTPIQVAAVSAVNGVLYEPYAKDWDPYTGCVY
 ESSAPKMKRQIISBETSKVRYALESVWARGTQCGAFVGYRVGGKTGTAOKAGRY
 LNNHIVSFIQFADDPQIIVVVAIDNPKDTVQFGVVAAPIVGKIIGDSIQALGIE
 KRNQIEKERKNDPELVEVDLVGRTRKDLHESYYELKIDADGKGVVVAQAPERT
 KVAVGSTIRLYMDDTPTD"
 /complement(2295..4466)
 /gene="pbbp"
 /complement(2295..4466)
 /gene="pbbp"
 /note="BH2573"
 /codon_start=1
 /transl_table=11
 /product="penicillin-binding protein 2B (cell-division septum)"
 /protein_id="BAB06292.1"
 /db_xref="GI:10175194"
 /translation="MDIKRAATNKRAMLVLFIVVFSVLIAFLYLQIEKVKYSL
 QAIABARTSVSEVLHKGRTIRDRGDALAEVTSVVAVLNDSHOPAVEPRTAQ
 LLAHINNSERLLELTDRVQVGLGAGANLTQQRDKIADLDLPGIYFTEERY
 YPKHYASHLIGITDRDMGSRGLESDEYLRGDSRIRFKQCGIIPRPSERI
 EPKATGNDVYITLDSNIQTALDQVMTQVEEYKPERIIAIVADPKTQGLIAMSPTF
 NPNEYQITNTMYAVSDRPGSTMKVFTLAAALEEVLNVEQYQSGTYAIRNDTN
 PPIRDHQCRGWGITISYLEAMQSSNVGFSIALEKLGPEKLYLIDFEGFGEITD
 LPNEAASVYKALRDAASTAFQGTAVTPIQIQIQAATAIANDGKMKPVVVDRTVDS
 ETGEVIEKEPEVVGEPISEKETAKEVRDILETVTVTSSSGTGRPFYLEGPDVAGTGA

OVNRDPGCVLNGHCKNLESPFGMAPKDDPSVIVVAVDRSLNTNOVSESPVAKIEN
 TIMNSLOVLINISPSVEHLKDEEGYELADPAGESARSAREELEQAGMKYVILGEGD
 TVGGQPSYGHKLLGERVILATESYTLPSMIGWSDVLKVNVLVNVNVLFGQG
 FVVGSIIEGDTITPDYVVELASPTKRKTXTDEKDEE"
 /complement(4529..4891)
 /gene="BH2574"
 /complement(4529..4891)
 /note="BH2574"
 /unknown conserved protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06293.1"
 /db_xref="GI:10175195"
 /translation="MCNARSIEHVRRQPAQPOAPRPRRQCKITKGEKMLGLM
 TVAPLLFIFGSVNYATMVSVNRHELEQTVAAEQINDGLSLQVVELSSPERILKI
 AKSLGKJEDENKVVQN"
 /complement(4939..5871)
 /gene="BH2575"
 /complement(4939..5871)
 /note="BH2575"
 /unknown conserved protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06294.1"
 /db_xref="GI:10175196"
 /translation="MPEHVTLVKNSVIGLNKPDGIYVDCITLGGAGHSQEVKOLTG
 GGHVAFDODQYALSHAKETLSPYSGFTLIESNFRYMKLEELGVHVDGLVFLDG
 VSPQLDEDERGFSYHEDAPLDMNRQSLSAYEVNQWDFIELMKILISRYGERFPA
 KQAKIEQSRKKRTITGELVDIIKEAPAPARTGTHGPAKRTFQAIRAVNDELG
 AFEEALEAIDLTPAKRIVCVITFHSLEDRIKCKEMPRESKGPDPVPPGLVIPESYKA
 TLKLTIKDPIVFTAEIILNRRARSAKLRIAEQ"
 /complement(5901..6332)
 /gene="BH2576"
 /complement(5901..6332)
 /note="BH2576"
 /unknown conserved protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06295.1"
 /db_xref="GI:10175197"
 /translation="MPMGYRNNDKGMIIKAKREBELGETFVVTGRLDRCLPVYP
 QVWKKLRESNLPTKADAPRTFFSGCELCEDKQGRVNIASIPREPAQLKKE
 CVVIGSVNRVETWSKELMBEYFAESESEFSAENIVDFDL"
 /complement(6487..8103)
 /gene="BH2577"
 /complement(6487..8103)
 /note="BH2577"
 /unknown conserved protein in B. subtilis"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB06296.1"
 /db_xref="GI:10175198"
 /translation="MIVBELKLLPSKAAKQYLNNDMLSPFDYNIHOPTVFOORLS
 DLQOPYDRDALSALLSYQKFAFHDKAAQVVKLKDPRSVVIVGGQAGLLTGPLY
 TIYKATVILAREQERALGVVVPVFWIAGEDHLDREINAVPIEKNGRWSHRIEEK
 RKXIAEAGLNKETLAKWLATVRSLPETHEHLPLYERVKTLAGNSNTYTFPABELL

Query Match 6.9%; Score 52; DB 1; Length 300950;
 Best Local Similarity 44.2%; Pred. No. 0.00035;
 Matches 214; Conservative 0; Mismatches 270; Indels 0; Gaps 0;
 Qy 32 TCGTTCTTCGCGAGGAGCTGCAAGGCGATAGCCACATAGTGTGTTTGAAGCTATAA 91
 Db 16087 TGGCTTTAGGCTCAGTGGAGCTCGTGGGTATGCATATCGTGTGTTGAAGATTATG 16028
 Qy 92 ACCAGCTCGGTAAAGCGTGGAGGCTTTTAAGCGGGGTAGCGCCGGGGAATCGTTTCGG 151
 Db 16027 AACAGGAAAGATACCGCATAGACTATCTCCCGGTAGTAGTATGGGAGCCCTTGTGGCCT 15968

[illegible]

JOURNAL Gene 258 (1-2), 127-139 (2000)
MEDLINE 20564182
REFERENCES 4 (sites)
AUTHORS Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasaki, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and Shinagawa, H.
TITLE Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12
JOURNAL DNA Res. 8 (1), 11-22 (2001)
MEDLINE 21156231
REFERENCES 5 (bases 1 to 291136)
AUTHORS Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and Hayashi, T.
TITLE Direct Submission
JOURNAL Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: kene@gen-info.osaka-u.ac.jp, URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365, Fax: 81-6-6879-2047)
COMMENT genome project.
FEATURES
source location/Qualifiers
1. 291136
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/db_strain="RIMD 0509952"
/db_xref="taxon:83334"
126. .380
/gene="ECol1692"
126. .380
/gene="ECol1692"
/note="similar to YMG_ECOLI gi|1787445 percent identity 97 in 84 aa (Conserved in E. coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB35115.1"
/db_xref="GI:13361157"
/translation="MGITAMITFGLIAGIANKLIPGRDGGFPLTCLIGVAVNGS WATMFGISGISGFNLHSPFLVAVGALIVLGFLLRRE"
complement (430. .2400)
/gene="ECol1693"
complement (430. .2400)
/gene="ECol1693"
/note="probable TonB dependent outer membrane receptor, similar to TonB dependent outer membrane receptor PirA [Escherichia coli CFT073] gi|3661477|gb|AAC61709.1| percent identity 97 in 656 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative TonB dependent outer membrane receptor"
/protein_id="BAB35116.1"
/db_xref="GI:13361158"
/translation="MKLKHYLCTALSLAFTQAAVAQESDLTWSSPVSSTTTVL DQPTMALDKONVAOLSVPGVLOKSGSRBQVKGFPDSQVPPVPGVPIYV YQENLDLARIITNNLGNVSKGYSILQGPOMQGAINTQKPTKPLBASIRBQG WSRSDNARIDMAHSPASSRIGYLSGSLQODDPLGPHGVNNDLAKHGMINSSA DDKGQVKGAPFTRENDVTLTYTKDQKDNPPYSGNSGQSRVQWNPBPKESFY QGTTQLANDRFTLSRLYRDPENTMLNYSVLADLNKKSYSYSDYSQALQALAD DWRSDEAKKHEKDGSIITYDDNNOCAFVWQWKGKHFVNEBTLALSYDRKRPFTLK ERYTTSKPAVNOIAIVNPKLPERARGVLTWNGAFTHDMGFEVYVYRVDALISH NIDADTIONSGTVDYSGLDAGIKGKICNIIIDVGLYLIHADAKRIDGTTDLPT QMTGAMTLLKMPRLSVTLSEPARSSVSNSGSOQAAGFAFTHIRAYTTLGHGSVN ASVNNLEFDIYRAISBGFVSEGRNFMAGVETTP"
complement (2426. .3280)
/gene="ECol1694"
complement (2426. .3280)
/gene="ECol1694"
/note="molybdenum transport protein, similar to molybdenum
transport proteins e.g. gi|3661478|gb|AAC61710.1| percent identity 91 in 284 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="molybdenum transport protein"
/protein_id="BAB35117.1"
/db_xref="GI:13361159"
/translation="MIFLSQAOIDALLIEDIQGDDLTTFALNTIGHQGYIEFLRQCG CVSGVAVACKMLTTLGITIDVADSGQANAGDRLIAQGNAAALHQGWALQNVLEW SCGVSDYDQMLALIERYPDGNACTRKAIPGTRILASQALIAAGLIRACGAETI ILFANRHFLEHNDQMSGAINQLRRAHPEKTIYEADTFPKAIIAAGDVQLODFE SPOQATRIAOIAPSLAPHCITLALGGINLTLLKNYLDGCGRLRFTPSAPYAADADIKV SIQPNAST"
complement (3219. .3500)
/gene="ECol1695"
complement (3219. .3500)
/gene="ECol1695"
/note="unknown"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB35118.1"
/db_xref="GI:13361160"
/translation="MPTLISVDKIARITAPRNVLSRTSAGVLARLTWSVSGYIAG INNRVLPSPLPAARGSSGIIAYRRRHCDPFPSSGNNRCVTSRGYRGR"
complement (3445. .4089)
/gene="ECol1696"
complement (3445. .4089)
/gene="ECol1696"
/note="unknown, similar to hypothetical proteins e.g. ORF2 [Escherichia coli CFT073] gi|3661479|gb|AAC61711.1| percent identity 98 in 214 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB35119.1"
/db_xref="GI:13361161"
/translation="MLTIDIPRADIYLOQLTLARTKTEPDMRQARAKAENCASPT BSYLOQLAKTIDLOQAGQTLFDMGCGPQVSLADKLTITIGVYSQGLNVAARRAA AKADNVHLIORAMEBWSDLPRCDIAVASSSTLVADRRQMSLNQARLRVYTHL VSTSFVSPAIQRAAGREVILEPNYIFALNVLYQWGIYAHVDIFRQNCODNSTY"
complement (4099. .4857)
/gene="ECol1697"
complement (4099. .4857)
/gene="ECol1697"
/note="probable ferric enterobactin transport, similar to ferric enterobactin transport ATP-binding protein [Escherichia coli CFT073] gi|3661480|gb|AAC61712.1| percent identity 79 in 148 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative ferric enterobactin transport ATP-binding protein"
/protein_id="BAB35120.1"
/db_xref="GI:13361162"
/translation="MTLALRHASLGSYRHPVLPADVSTLSOGTICCLLGANGCKT LMSLITGVPILKGVLDLSLPVOTLSHRQAAQAIAPWROAGIIPASVYDMLMGL APTLGAISVPGQERIKATBOLEKGIITHLAARPNTISGGRQALALRALVQAPRL LILDERPASSLDLSDHOIQLDITDAQLKNNGMTLMSTHPLAHNATDSITQVEPGRV TQQLPTEQTLTKLAALYVNSADQIHNLISLISH"
complement (4854. .5834)
/gene="ECol1698"
complement (4854. .5834)
/gene="ECol1698"
/note="probable ABC transporter protein (permease), similar to ABC transporter permeases e.g. [Haemophilus influenzae] gi|2501391|sp|O57130|YE71_HAEIN percent identity 40 in 323 aa"
/codon_start=1
transport proteins e.g. gi|3661478|gb|AAC61710.1| percent identity 91 in 284 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="molybdenum transport protein"
/protein_id="BAB35117.1"
/db_xref="GI:13361159"
/translation="MIFLSQAOIDALLIEDIQGDDLTTFALNTIGHQGYIEFLRQCG CVSGVAVACKMLTTLGITIDVADSGQANAGDRLIAQGNAAALHQGWALQNVLEW SCGVSDYDQMLALIERYPDGNACTRKAIPGTRILASQALIAAGLIRACGAETI ILFANRHFLEHNDQMSGAINQLRRAHPEKTIYEADTFPKAIIAAGDVQLODFE SPOQATRIAOIAPSLAPHCITLALGGINLTLLKNYLDGCGRLRFTPSAPYAADADIKV SIQPNAST"
complement (3219. .3500)
/gene="ECol1695"
complement (3219. .3500)
/gene="ECol1695"
/note="unknown"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB35118.1"
/db_xref="GI:13361160"
/translation="MPTLISVDKIARITAPRNVLSRTSAGVLARLTWSVSGYIAG INNRVLPSPLPAARGSSGIIAYRRRHCDPFPSSGNNRCVTSRGYRGR"
complement (3445. .4089)
/gene="ECol1696"
complement (3445. .4089)
/gene="ECol1696"
/note="unknown, similar to hypothetical proteins e.g. ORF2 [Escherichia coli CFT073] gi|3661479|gb|AAC61711.1| percent identity 98 in 214 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB35119.1"
/db_xref="GI:13361161"
/translation="MLTIDIPRADIYLOQLTLARTKTEPDMRQARAKAENCASPT BSYLOQLAKTIDLOQAGQTLFDMGCGPQVSLADKLTITIGVYSQGLNVAARRAA AKADNVHLIORAMEBWSDLPRCDIAVASSSTLVADRRQMSLNQARLRVYTHL VSTSFVSPAIQRAAGREVILEPNYIFALNVLYQWGIYAHVDIFRQNCODNSTY"
complement (4099. .4857)
/gene="ECol1697"
complement (4099. .4857)
/gene="ECol1697"
/note="probable ferric enterobactin transport, similar to ferric enterobactin transport ATP-binding protein [Escherichia coli CFT073] gi|3661480|gb|AAC61712.1| percent identity 79 in 148 aa"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative ferric enterobactin transport ATP-binding protein"
/protein_id="BAB35120.1"
/db_xref="GI:13361162"
/translation="MTLALRHASLGSYRHPVLPADVSTLSOGTICCLLGANGCKT LMSLITGVPILKGVLDLSLPVOTLSHRQAAQAIAPWROAGIIPASVYDMLMGL APTLGAISVPGQERIKATBOLEKGIITHLAARPNTISGGRQALALRALVQAPRL LILDERPASSLDLSDHOIQLDITDAQLKNNGMTLMSTHPLAHNATDSITQVEPGRV TQQLPTEQTLTKLAALYVNSADQIHNLISLISH"
complement (4854. .5834)
/gene="ECol1698"
complement (4854. .5834)
/gene="ECol1698"
/note="probable ABC transporter protein (permease), similar to ABC transporter permeases e.g. [Haemophilus influenzae] gi|2501391|sp|O57130|YE71_HAEIN percent identity 40 in 323 aa"
/codon_start=1


```

/evidence-not experimental
/transl_table=11
/product="putative ABC transporter permease protein"
/protein_id="BAB35121.1"
/db_xref="GI:13361163"
/translation="MRIVNGCILLAAISITPAANSYAHYLDMOQLLALILROBNFV
QEQIVFQWRVPRILAAALLGAALAGAGTTTCYQMLRNPLVSPDILGVSAAGLGACTA
ILWGLSVITQIYAFCCGGLMVVAGVWILTRVRHDPILTLVLVGLAGTLCGAGISL
IKTLADPYTQLPSITFWLGGSLSTVLRDLCYAAPILIGLSLPLFLWRNMLLTLSD
DEARSLGNVLRFLVLCATLITASTVAIGLIGWGLVVPVHIALTLTGHNQQLL
PMACVCAILLTLTDTLARSIGTTEIPILGILTAFCVAPFFLLLLLRGRQ"
/gene="Ece1699"
/complement(5834..6856)
/genes="Ece1699"
/notes="probable ABC transporter, weakly similar to iron
binding protein [Pyrococcus
(iii) ABC transporter, ATP-binding protein [Pyrococcus
Query Match 6.4%; Score 49.8; DB 1; Length 291136;
Best Local Similarity 52.7%; Pred. No. 0.0016;
Matches 108; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
QY 340 GCGACGATTTATCTCGGAGGGCTCTATACCTCTCGGAGGAGTTTAAATCCCGCA 399
DB 47183 GCCACCAATTTAGTACTGGCGTGAATATGTTTACTGAGGCGATCTCCATCTTGT 47242
QY 400 CTCTCGCAGCTGTGCANTCCCGGCATATTGTAACCCGTTGAGTATAGAAATPACTTG 459
DB 47243 ATTCCGCAATCATGCGATATTCCAGGACTAATGCGACCCGTTGCGCATACGGCTACTGG 47302
QY 460 CTCTGTCAGGAGTATAGTTTAAACCTTCCGTTGAGCCCTTTTCAGGAAAGCGTATT 519
DB 47303 CTGTTGATGCGGAGTCTGTAAACCAATCTCTATTCGCTCACGCGTGCATGGTGTCT 47362
QY 520 CCCACCGTTTGCCTGATGCTCTTC 544
DB 47363 GATATTGTGATAGCGGTGACCTGC 47387

```

```

RESULT 15
ECOTGSA/C ECOTGSA 3653 bp DNA linear BCT 18-MAR-1994
LOCUS Escherichia coli tgs gene, partial cds.
DEFINITION M64675
ACCESSION M64675
VERSION M64675.1 GI:402693
KEYWORDS tgs gene.
SOURCE Escherichia coli (strain K-12) DNA.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 3653)
AUTHORS Boel.M. and Kersten,H.
TITLE Organization and functions of genes in the upstream region of tyrT
of Escherichia coli: phenotypes of mutants with partial deletion of
a new gene (tgs)
JOURNAL J. Bacteriol. 176 (1), 221-231 (1994)
MEDLINE 94110230
PUBMED 8282700
FEATURES
source Location/Qualifiers
1..3653
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
1154..1159
TATA_signal
-35_signal 1176..1181
TATA_signal 2213..2218
gene 2524..3479
/gene="tgs"
-35_signal 2524..2529
/gene="tgs"
TATA_signal 2548..2553
/gene="tgs"
CDS <2687..3479

```

```

/gene="tgs"
/codon_start=2
/transl_table=11
/protein_id="AAA16860.1"
/db_xref="GI:402694"
/translation="GLIARITNICYKHELNIVQNEFVDHRTGRFMRTELEGI FND
TLLADLSALPEGSVRELNPARRRIIVLTKEAHCGLDLMKANVGLDVBIAAVIG
NHDTRLSVERFDIPPELVSHGELTRNEHQKMAIDAQPDYVVLAKYMRVLTPEF
VARPNKTIINIHSLFPAFICARPYHOAYERGKVICATARYNDNLDEGPIIMQDVI
HYDHTYTAEDMMRAGROVEKNVLSRALYKVLQAORVYVYGNRTLL"
BASE COUNT 948 a 916 c 843 g 946 t
ORIGIN
Query Match 6.4%; Score 48.2; DB 1; Length 3653;
Best Local Similarity 52.2%; Pred. No. 0.0041;
Matches 107; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
QY 340 GCGACGATTTATCTCGGAGGGCTCTATACCTCTCGGAGGAGTTTAAATCCCGCA 399
DB 1680 GCCACCAATTTAGTACCGGACGTAATATGTTTACTGAGGCGATCTCCATCTTGT 1621
QY 400 CTCTCGGAGCTGTGCATATCCCGGCATATTTGAACCCGTTGAGTATAGAAATPACTTG 459
DB 1620 ATTCCGCAATCATGCAATATCCAGGACTCATGCACTGTTGCATACGGCTACTGG 1561
QY 460 CTCTGTTGAGGAGGTATAGTTTAAACCTTCCCGTTGAGCCCTTTTCAGGAAAGCGTATT 519
DB 1560 CTGTTGATGAGGAGCTGTTTAAACCAATCTCTATTTCCCTCACGCGTGCATTTGGGGGCT 1501
QY 520 CCCACCGTTTGCCTGATGCTCTTC 544
DB 1500 GATATTGTGATAGCGGTGACCTGC 1476

```

Search completed: June 19, 2003, 14:02:48
Job time : 1454 secs